(19) World Intellectual Property Organization International Bureau





(43) International Publication Date 1 August 2002 (01.08.2002)

PCT

(10) International Publication Number WO 02/059321 A2

(51) International Patent Classification⁷: 15/51, 5/10, C07K 14/18, C12N 15/85

C12N 15/40,

(21) International Application Number: PCT/EP02/00526

(22) International Filing Date: 16 January 2002 (16.01.2002)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 60/263,479 23 January 2001 (23.01.2001) U

. (71) Applicant (for all designated States except US): ISTITUTO DI RICERCHE DI BIOLOGIA MOLECO-LARE P. ANGELETTI SPA [IT/IT]; Via Pontina Km. 30, 600, 1-00040 Pomezia (Rome) (IT).

(72) Inventors; and

(75) Inventors/Applicants (for US only): DE FRANCESCO, Raffaele [IT/IT]; IRBM, Via Pontina Km. 30, 600, I-00040 Pomezia (Rome) (IT). MIGLIACCIO, Giovanni [IT/IT]; IRBM, Via Pontina Km. 30, 600, I-00040 Pomezia (Rome) (IT). PAONESSA, Giacomo [IT/IT]; IRBM, Via Pontina Km. 30, 600, I-00040 Pomezia (Rome) (IT).

(74) Agent: THOMPSON, John; Merck & Co., Inc., European Patent Department, Terlings Park, Eastwick Road, Harlow, Essex CM20 2QR (GB).

(81) Designated States (national): CA, JP, US.

(84) Designated States (regional): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:

 without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



TITLE OF THE INVENTION HEPATITIS C VIRUS REPLICONS AND REPLICON ENHANCED CELLS

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims priority to U.S. Serial No. 60/263,479, filed January 23, 2001, hereby incorporated by reference herein.

BACKGROUND OF THE INVENTION

5

15

20

25

30

35

The references cited in the present application are not admitted to be prior art to the claimed invention.

It is estimated that about 3% of the world's population are infected with the Hepatitis C virus (HCV). (Wasley, et al., 2000. Semin. Liver Dis. 20, 1-16.) Exposure to HCV results in an overt acute disease in a small percentage of cases, while in most instances the virus establishes a chronic infection causing liver inflammation and slowly progresses into liver failure and cirrhosis. (Iwarson, 1994. FEMS Microbiol. Rev. 14, 201-204.) In addition, epidemiological surveys indicate an important role of HCV in the pathogenesis of hepatocellular carcinoma. (Kew, 1994. FEMS Microbiol. Rev. 14, 211-220, Alter, 1995. Blood 85, 1681-1695.)

The HCV genome consists of a single strand RNA of about 9.5 kb in length, encoding a precursor polyprotein of about 3000 amino acids. (Choo, et al., 1989. Science 244, 362-364, Choo, et al., 1989. Science 244, 359-362, Takamizawa, et al., 1991. J. Virol. 65, 1105-1113.) The HCV polyprotein contains the viral proteins in the order: C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B.

Individual viral proteins are produced by proteolysis of the HCV polyprotein. Host cell proteases release the putative structural proteins C, E1, E2, and p7, and create the N-terminus of NS2 at amino acid 810. (Mizushima, et al., 1994. J. Virol. 68, 2731-2734, Hijikata, et al., 1993. P.N.A.S. USA 90, 10773-10777.)

The non-structural proteins NS3, NS4A, NS4B, NS5A and NS5B presumably form the virus replication machinery and are released from the polyprotein. A zinc-dependent protease associated with NS2 and the N-terminus of NS3 is responsible for cleavage between NS2 and NS3. (Grakoui, et al., 1993. J. Virol. 67, 1385-1395, Hijikata, et al., 1993. P.N.A.S. USA 90, 10773-10777.) A distinct serine protease located in the N-terminal domain of NS3 is responsible for proteolytic cleavages at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/NS5B junctions. (Barthenschlager, et al., 1993. J. Virol. 67, 3835-3844, Grakoui, et al.,

1993. Proc. Natl. Acad. Sci. USA 90, 10583-10587, Tomei, et al., 1993. J. Virol. 67, 4017-4026.) NS4A provides a cofactor for NS3 activity. (Failla, et al., J. Virol. 1994. 68, 3753-3760, De Francesco, et al., U.S. Patent No. 5,739,002.) NS5A is a highly phosphorylated protein concurring interferon resistance. (De Francesco, et al., 2000. Semin Liver Dis., 20(1), 69-83, Pawlotsky, 1999. J. Viral Hepat. Suppl. 1, 47-48.) NS5B provides an RNA polymerase. (De Francesco, et al., International Publication Number WO 96/37619, Behrens, et al., 1996. EMBO 15, 12-22, Lohmann, et al., 1998. Virology 249, 108-118.)

Lohmann, et al., Science 285, 110-113, 1999, illustrates the ability of a biscistronic HCV replicon to replicate in a hepatoma cell line. The biscistonic HCV replicon contained a neomycin cistron and an NS2-NS5B or an NS3-NS5B cistron. "NS2-NS5B" refers to a NS2-NS3-NS4A-NS4B-NS5A-NS5B polyprotein. "NS3-NS5B" refers to a NS3-NS4A-NS4B-NS5A-NS5B polyprotein.

Bartenschlager, European Patent Application 1 043 399, published

October 11, 2000 (not admitted to be prior art to the claimed invention), describes a cell culture system for autonomous HCV RNA replication and protein expression.

Replication and protein expression is indicated to occur in sufficiently large amounts for quantitative determination. European Patent Application 1 043 399 indicates that prior cell lines or primary cell cultures infected with HCV do not provide favorable circumstances for detecting HCV replication.

SUMMARY OF THE INVENTION

25

30

35

The present invention features nucleic acid containing one or more adaptive mutations, and HCV replicon enhanced cells. Adaptive mutations are mutations that enhance HCV replicon activity. HCV replicon enhanced cells are cells having an increased ability to maintain an HCV replicon.

An HCV replicon is an RNA molecule able to autonomously replicate in a cultured cell and produce detectable levels of one or more HCV proteins. The basic subunit of an HCV replicon encodes for a HCV NS3-NS5B polyprotein along with a suitable 5' UTR-partial core (PC) region and 3' UTR. The 5' UTR-PC region is made up of a 5'UTR region and about 36 nucleotides of the beginning of the core. Additional regions may be present including those coding for HCV proteins or elements such as the complete core, E1, E2, p7 or NS2; and those coding for other types of proteins or elements such as a encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES), a reporter protein or a selection protein.

The present application identifies different adaptive mutations that enhance HCV replicon activity. Enhancing replicon activity brings about at least one of the following: an increase in replicon maintenance in a cell, an increase in replicon replication, and an increase in replicon protein expression.

5

10

15

20

25

30

35

Adaptive mutations are described herein by identifying the location of the adaptive mutation with respect to a reference sequence present in a particular region. Based on the provided reference sequence, the same adaptive mutation can be produced in corresponding locations of equivalent regions having an amino acid sequence different than the reference sequence. Equivalent regions have the same function or encode for a polypeptide having the same function.

Replicon enhanced cells are a preferred host for the insertion and expression of an HCV replicon. Replicon enhanced cells are initially produced by creating a cell containing a HCV replicon and then curing the cell of the replicon. The term "replicon enhanced cell" includes cells cured of HCV replicons and progeny of such cells.

Thus, a first aspect of the present invention describes a nucleic acid molecule comprising at least one of the following regions: an altered NS3 encoding region, an altered NS5A encoding region, and an altered EMCV IRES region. The altered region contains one or more adaptive mutations. Reference to the presence of particular adaptive mutation(s) does not exclude other mutations or adaptive mutations from being present. Adaptive mutations are described with reference to either an encoded amino acid sequence or a nucleic acid sequence.

A nucleic acid molecule can be single-stranded or part of a double strand, and can be RNA or DNA. Depending upon the structure of the nucleic acid molecule, the molecule may be used as a replicon or in the production of a replicon. For example, single-stranded RNA having the proper regions can be a replicon, while double-stranded DNA that includes the complement of a sequence coding for a replicon or replicon intermediate may useful in the production of the replicon or replicon intermediate.

Preferred nucleic acid molecules are those containing region(s) from SEQ. ID. NOs. 1, 2, or 3, or the RNA version thereof, with one or more adaptive mutations. Reference to "the RNA version thereof" indicates a ribose backbone and the presence of uracil instead of thymine.

The presence of a region containing an adaptive mutation indicates that at least one such region is present. In different embodiments, for example, adaptive

mutations described herein are present at least in the NS3 region, in the NS5A region, in the NS3 and NS5A regions, in the EMCV IRES and NS3 regions, in the EMCV and NS5A regions, and in the ECMV IRES, NS3 and NS5A regions.

Another aspect of the present invention describes an expression vector comprising a nucleotide sequence of an HCV replicon or replicon intermediate coupled to an exogenous promoter. Reference to a nucleotide sequence "coupled to an exogenous promoter" indicates the presence and positioning of an RNA promoter such that it can mediate transcription of the nucleotide sequence and that the promoter is not naturally associated with the nucleotide sequence being transcribed. The expression vector can be used to produce RNA replicons.

5

10

15

20

25

30

35

Another aspect of the present invention describes a recombinant human hepatoma cell. Reference to a recombinant cell includes an initially produced cell and progeny thereof.

Another aspect of the present invention describes a method of making a HCV replicon enhanced cell. The method involves the steps of: (a) introducing and maintaining an HCV replicon into a cell and (b) curing the cell of the HCV replicon.

Another aspect of the present invention describes an HCV replicon enhanced cell made by a process comprising the steps of: (a) introducing and maintaining an HCV replicon into a cell and (b) curing the cell of the HCV replicon.

Another aspect of the present invention describes a method of making a HCV replicon enhanced cell comprising an HCV replicon. The method involves (a) introducing and maintaining a first HCV replicon into a cell, (b) curing the cell of the replicon, and (c) introducing and maintaining a second replicon into the cured cell, where the second replicon may be the same or different as the first replicon.

Another aspect of the present invention describes an HCV replicon enhanced cell containing a HCV replicon made by the process involving the step of introducing an HCV replicon into an HCV replicon enhanced cell. The HCV replicon introduced into the HCV replicon enhanced cell may be the same or different than the HCV replicon used to produce the HCV replicon enhanced cell. In a preferred embodiment, the HCV replicon introduced into an HCV replicon enhanced cell is the same replicon as was used to produce the enhanced cell.

Another aspect of the present invention describes a method of measuring the ability of a compound to affect HCV activity using an HCV replicon comprising an adaptive mutation described herein. The method involves providing a compound to a cell comprising the HCV replicon and measuring the ability of the

compound to affect one or more replicon activities as a measure of the effect on HCV activity.

Another aspect of the present invention describes a method of measuring the ability of a compound to affect HCV activity using an HCV replicon enhanced cell that comprises an HCV replicon. The method involves providing a compound to the cell and measuring the ability of the compound to effect one or more replicon activities as a measure of the effect on HCV activity.

Other features and advantages of the present invention are apparent from the additional descriptions provided herein including the different examples. The provided examples illustrate different components and methodology useful in practicing the present invention. The examples do not limit the claimed invention. Based on the present disclosure the skilled artisan can identify and employ other components and methodology useful for practicing the present invention.

15 BRIEF DESCRIPTION OF THE DRAWINGS

5

10

Figures 1A-1G illustrate the nucleic acid sequence for the pHCVNeo.17 coding strand (SEQ. ID. NO. 3). The different regions of pHCVNeo.17 are provided as follows:

- 1-341: HCV 5' non-translated region, drives translation of the core-neo fusion protein;
- 20 342-1181: Core-neo fusion protein, selectable marker;
 - 1190-1800: Internal ribosome entry site of the encephalomyocarditis virus, drives translation of the HCV NS region;
 - 1801-7755: HCV polyprotein from non-structural protein 3 to non-structural protein 5B;
- 25 1801-3696: Non-structural protein 3 (NS3), HCV NS3 protease/helicase;
 - 3697-3858: Non-structural protein 4A (NS4A), NS3 protease cofactor;
 - 3859-4641: Non-structural protein 4B (NS4B);
 - 4642-5982: Non-structural protein 5A (NS5A);
 - 5983-7755: Non-structural protein 5B (NS5B); RNA-dependent RNA polymerase
- 30 7759-7989: HCV 3' non-translated region; and
 - 7990-10690 plasmid sequences comprising origin of replication, beta lactamase coding sequence, and T7 promoter.

DETAILED DESCRIPTION OF THE INVENTION

5

10

15

20

25

30

35

HCV replicons and HCV replicon enhanced cells can be used to produce a cell culture providing detectable levels of HCV RNA and HCV protein. HCV replicons and HCV replicon enhanced hosts can both be obtained by selecting for the ability to maintain an HCV replicon in a cell. As illustrated in the examples provided below, adaptive mutations present in HCV replicons and host cells can both assist replicon maintenance in a cell.

The detectable replication and expression of HCV RNA in a cell culture system has a variety of different uses including being used to study HCV replication and expression, to study HCV and host cell interactions, to produce HCV RNA, to produce HCV proteins, and to provide a system for measuring the ability of a compound to modulate one or more HCV activities.

Preferred cells for use with a HCV replicon are Huh-7 cells and Huh-7 derived cells. "Huh-7 derived cells" are cell produced starting with Huh-7 cells and introducing one or more phenotypic and/or genotypic modifications.

Adaptive Mutations

Adaptive mutations enhance the ability of an HCV replicon to be maintained and expressed in a host cell. Adaptive mutations can be initially selected for using a wild type HCV RNA construct or a mutated HCV replicon. Initial selection involves providing HCV replicons to cells and identifying clones containing a replicon.

Nucleic acid sequences of identified HCV replicons can be determined using standard sequencing techniques. Comparing the sequence of input HCV constructs and selected constructs provides the location of mutations. The effect of particular mutation(s) can be measured by, for example, producing a construct to contain particular mutation(s) and measuring the effect of these mutation(s). Suitable control constructs for comparison purposes include wild type constructs and constructs previously evaluated.

Adaptive mutations were predominantly found in the HCV NS3 and NS5A regions. With the exception of two silent mutations in NS5A and NS5B, consensus mutations occurring in the NS region resulted in changes to the deduced amino acid sequence. Noticeably, the amino acid changes occurred in residues that are conserved in all or a large number of natural HCV isolates. HCV sequences are well known in the art and can be found, for example, in GenBank.

Adaptive mutations described herein can be identified with respect to a reference sequence. The reference sequence provides the location of the adaptive mutation in, for example, the NS3 or NS5A RNA, cDNA, or amino acid sequence. The remainder of the sequence encodes for a functional protein that may have the same, or a different, sequence than the reference sequence.

Preferred NS3 and NS5A adaptive mutations and examples of changes that can be made to produce such mutations are shown in Tables 1 and 2. The amino acid numbering shown in Tables 1 and 2 is with respect to SEQ. ID. NO. 1. The nucleotide numbering shown in Tables 1 and 2 is with respect to SEQ. ID. NO. 2. SEQ. ID. NO. 1 provides the amino acid sequence of the Con1 HCV isolate (Accession Number AJ238799). SEQ. ID. NO. 2 provides the nucleic acid sequence of the Con1 HCV isolate.

TABLE 1

15

10

5

Preferred NS3 Adaptive Mutations					
Amino Acid	Nucleotide				
gly1095ala	G3625C				
glu1202gly	A3946G				
ala1347thr	G4380A				

TABLE 2

Preferred NS5A A	Preferred NS5A Adaptive Mutations						
Amino Acid	Nucleotide						
Lys@2039	AAA@6458						
asn2041thr	A6463C						
ser2173phe	C6859T						
ser2197phe	C6931T						
leu2198ser	T6934C						
ala2199thr	G6936A						
ser2204arg	C6953A (or G)						

[&]quot;@" refers to an addition.

Preferred adaptive mutations identified with respect to a reference sequence can be produced changing the encoding region of SEQ. ID. NO. 1, or an equivalent sequence, to result in the indicated change. Preferred adaptive mutations provided in Tables 1 and 2 occur in amino acids conserved among different HCV isolates.

Adaptive mutations have different effects. Some mutations alone, or in combination with other mutations, enhance HCV replicon activity. In some cases, two or more mutations led to synergistic effects and in one case, a slightly antagonistic effect was observed.

An adaptive mutation once identified can be introduced into a starting construct using standard genetic techniques. Examples of such techniques are provided by Ausubel, *Current Protocols in Molecular Biology*, John Wiley, 1987-1998, and Sambrook, *et al.*, *Molecular Cloning*, *A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, 1989.

HCV replicons containing adaptive mutations can be built around an NS3 region or NS5A region containing one or more adaptive mutations described herein. The final replicon will contain replicon components needed for replication and may contain additional components.

SEQ. ID. NO. 2 can be used as a reference point for different HCV regions as follows:

5' UTR- nucleotides 1-341;

Core- nucleotides 342-914;

E1- nucleotides 915-1490;

25 E2- nucleotides 1491-2579;

5

10

15

35

P7- nucleotides 2580-2768;

NS2- nucleotides 2769-3419;

NS3- nucleotides 3420-5312;

NS4A- nucleotides 5313-5474;

30 NS4B- nucleotides 5475-6257;

NS5A- nucleotides 6258-7598;

NS5B- nucleotides 7599-9371; and

3' UTR-nucleotides 9374-9605.

The amino acid sequences of the different structural and non-structural regions is provided by SEQ. ID. NO. 1.

Nucleic acid sequences encoding for a particular amino acid can be produced taking into account the degeneracy of the genetic code. The degeneracy of the genetic code arises because almost all amino acids are encoded for by different combinations of nucleotide triplets or "codons". The translation of a particular codon into a particular amino acid is well known in the art (see, e.g., Lewin GENES IV, p. 119, Oxford University Press, 1990). Amino acids are encoded for by RNA codons as follows:

A=Ala=Alanine: codons GCA, GCC, GCG, GCU

C=Cys=Cysteine: codons UGC, UGU

5

10 D=Asp=Aspartic acid: codons GAC, GAU

E=Glu=Glutamic acid: codons GAA, GAG

F=Phe=Phenylalanine: codons UUC, UUU

G=Gly=Glycine: codons GGA, GGC, GGG, GGU

H=His=Histidine: codons CAC, CAU

15 I=Ile=Isoleucine: codons AUA, AUC, AUU

K=Lys=Lysine: codons AAA, AAG

L=Leu=Leucine: codons UUA, UUG, CUA, CUC, CUG, CUU

M=Met=Methionine: codon AUG

N=Asn=Asparagine: codons AAC, AAU

20 P=Pro=Proline: codons CCA, CCC, CCG, CCU

O=Gln=Glutamine: codons CAA, CAG

R=Arg=Arginine: codons AGA, AGG, CGA, CGC, CGG, CGU S=Ser=Serine: codons AGC, AGU, UCA, UCC, UCG, UCU

T=Thr=Threonine: codons ACA, ACC, ACG, ACU

25 V=Val=Valine: codons GUA, GUC, GUG, GUU

W=Trp=Tryptophan: codon UGG

Y=Tyr=Tyrosine: codons UAC, UAU.

Constructs, including subgenomic and genomic replicons, containing one or more of the adaptive mutations described herein can also contain additional mutations. The additional mutations may be adaptive mutations and mutations not substantially inhibiting replicon activity. Mutations not substantially inhibiting replicon activity provide for a replicon that can be introduced into a cell and have detectable activity.

HCV Replicon

HCV replicons include the full length HCV genome and subgenomic constructs. A basic HCV replicon is a subgenomic construct containing an HCV 5' UTR- PC region, an HCV NS3-NS5B polyprotein encoding region, and a HCV 3' UTR. Other nucleic acid regions can be present such as those providing for HCV NS2, structural HCV protein(s) and non-HCV sequences.

5

10

15

20

25

30

The HCV 5' UTR-PC region provides an internal ribosome entry site (IRES) for protein translation and elements needed for replication. The HCV 5'UTR-PC region includes naturally occurring HCV 5' UTR extending about 36 nucleotides into a HCV core encoding region, and functional derivatives thereof. The 5'-UTR-PC region can be present in different locations such as site downstream from a sequence encoding a selection protein, a reporter, protein, or an HCV polyprotein.

Functional derivatives of the 5'-UTR-PC region able to initiate translation and assist replication can be designed taking into structural requirements for HCV translation initiation. (See, for example, Honda, et al., 1996. Virology 222, 31-42). The affect of different modifications to a 5' UTR-PC region can be determined using techniques that measure replicon activity.

In addition to the HCV 5' UTR-PC region, non-HCV IRES elements can also be present in the replicon. The non-HCV IRES elements can be present in different locations including immediately upstream the region encoding for an HCV polyprotein. Examples of non-HCV IRES elements that can be used are the EMCV IRES, poliovirus IRES, and bovine viral diarrhea virus IRES.

The HCV 3' UTR assists HCV replication. HCV 3' UTR includes naturally occurring HCV 3' UTR and functional derivatives thereof. Naturally occurring 3' UTR's include a poly U tract and an additional region of about 100 nucleotides. (Tanaka, et al., 1996. J. Virol. 70, 3307-3312, Kolykhalov, et al., 1996. J. Virol. 70, 3363-3371.) At least in vivo, the 3' UTR appears to be essential for replication. (Kolykhalov, et al., 2000. J. Virol. 2000 4, 2046-2051.) Examples of naturally occurring 3' UTR derivatives are described by Bartenschlager International Publication Number EP 1 043 399.

The NS3-NS5B polyprotein encoding region provides for a polyprotein that can be processed in a cell into different proteins. Suitable NS3-NS5B polyprotein sequences that may be part of a replicon include those present in different HCV strains and functional equivalents thereof resulting in the processing of NS3-NS5B to

a produce functional replication machinery. Proper processing can be measured for by assaying, for example, NS5B RNA dependent RNA polymerase.

The ability of an NS5B protein to provide RNA polymerase activity can be measured using techniques well known in the art. (See, for example, De Franscesco, et al., International Publication Number WO 96/37619, Behrens, et al., 1996. EMBO 15:12-22, Lohmann, et al., 1998. Virology 249:108-118.) Preferably, the sequence of the active NS5B is substantially similar as that provided in SEQ. ID. NO. 1, or a wild type NS5B such as strains HCV-1, HCV-2, HCV-BK, HCV-J, HCV-N, HCV-H. A substantially similar sequence provides detectable HCV polymerase activity and contains 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 amino acid alterations to that present in a HCV NS5B polymerase. Preferably, no more than 1, 2, 3, 4 or 5 alterations are present.

Alterations to an amino acid sequence provide for substitution(s), insertion(s), deletion(s) or a combination thereof. Sites of different alterations can be designed taking into account the amino acid sequences of different NS5B polymerases to identify conserved and variable amino acid, and can be empirically determined.

HCV replicons can be produced in a wide variety of different cells and *in vitro*. Suitable cells allow for the transcription of a nucleic acid encoding for an HCV replicon.

20

25

30

35

15

5

10

Additional Sequences

An HCV replicon may contain non-HCV sequences in addition to HCV sequences. The additional sequences should not prevent replication and expression, and preferably serve a useful function. Sequences that can be used to serve a useful function include a selection sequence, a reporter sequence, transcription elements and translation elements.

Selection Sequence

A selection sequence in an HCV replicon facilitates the identification of a cell containing the replicon. Selection sequences are typically used in conjunction with some selective pressure that inhibits growth of cells not containing the selection sequence. Examples of selection sequences include sequences encoding for antibiotic resistance and ribozymes.

Antibiotic resistance can be used in conjunction with an antibiotic to select for cells containing replicons. Examples of selection sequences providing for

antibiotic resistance are sequences encoding resistance to neomycin, hygromycin, puromycin, or zeocin.

A ribozyme serving as a selection sequence can be used in conjunction with an inhibitory nucleic acid molecule that prevents cellular growth. The ribozyme recognizes and cleaves the inhibitory nucleic acid.

Reporter Sequence

5

10

15

20

25

30

35

A reporter sequence can be used to detect replicon replication or protein expression. Preferred reporter proteins are enzymatic proteins whose presence can be detected by measuring product produced by the protein. Examples of reporter proteins include, luciferase, beta-lactamase, secretory alkaline phosphatase, beta-glucuronidase, green fluorescent protein and its derivatives. In addition, a reporter nucleic acid sequence can be used to provide a reference sequence that can be targeted by a complementary nucleic acid. Hybridization of the complementary nucleic acid to its target can be determined using standard techniques.

Additional Sequence Configuration

Additional non-HCV sequences are preferable 5' or 3' of an HCV replicon genome or subgenomic genome region. However, the additional sequences can be located within an HCV genome as long as the sequences do not prevent detectable replicon activity. If desired, additional sequences can be separated from the replicon by using a ribozyme recognition sequence in conjunction with a ribozyme.

Additional sequences can be part of the same cistron as the HCV polyprotein or can be a separate cistron. If part of the same cistron, the selection or reporter sequence coding for a protein should result in a product that is either active as a chimeric protein or is cleaved inside a cell so it is separated from HCV protein.

Selection and reporter sequences encoding for a protein when present as a separate cistron should be associated with elements needed for translation. Such elements include a 5' IRES.

Detection Methods

Methods for detecting replicon activity include those measuring the production or activity of replicon RNA and encoded for protein. Measuring includes qualitative and quantitative analysis.

Techniques suitable for measuring RNA production include those detecting the presence or activity of RNA. The presence of RNA can be detected using, for example, complementary hybridization probes or quantitative PCR. Techniques for measuring hybridization between complementary nucleic acid and quantitative PCR are well known in the art. (See for example, Ausubel, Current Protocols in Molecular Biology, John Wiley, 1987-1998, Sambrook, et al., Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, 1989, and U.S. Patent No. 5,731,148.)

RNA enzymatic activity can be provided to the replicon by using a ribozyme sequence. Ribozyme activity can be measured using techniques detecting the ability of the ribozyme to cleave a target sequence.

Techniques for measuring protein production include those detecting the presence or activity of a produced protein. The presence of a particular protein can be determined by, for example, immunological techniques. Protein activity can be measured based on the activity of an HCV protein or a reporter protein sequence.

Techniques for measuring HCV protein activity vary depending upon the protein that is measured. Techniques for measuring the activity of different nonstructural proteins such as NS2/3, NS3, and NS5B, are well known in the art. (See, for example, references provided in the Background of the Invention.)

Assays measuring replicon activity also include those detecting virion production from a replicon that produces a virion; and those detecting a cytopathic effect from a replicon producing proteins exerting such an effect. Cytopathic effects can be detected by assays suitable to measure cell viability.

Assays measuring replicon activity can be used to evaluate the ability of a compound to modulate HCV activities. Such assays can be carried out by providing one or more test compounds to a cell expressing an HCV replicon and measuring the effect of the compound on replicon activity. If a preparation containing more than one compound is found to modulate replicon activity, individual compounds or smaller groups of compounds can be tested to identify replicon active compounds.

Compounds identified as inhibiting HCV activity can be used to produce replicon enhanced cells and may be therapeutic compounds. The ability of a compound to serve as a therapeutic compound can be confirmed using animal models such as a chimpanzee to measure efficacy and toxicity.

5

10

15

20

25

30

Replicon Enhanced Host Cell

Replicon enhanced cells are initially produced by selecting for a cell able to maintain an HCV replicon and then curing the cell of the replicon. Cells produced in this fashion were found to have an increased ability to maintain a replicon upon subsequent HCV replicon transfection.

Initial transfection can be performed using a wild-type replicon or a replicon containing one or more adaptive mutations. If a wild-type replicon is employed, the replicon should contain a selection sequence to facilitate replicon maintenance.

Cells can be cured of replicons using different techniques such as those employing replicon inhibitory agent. In addition, replication of HCV replicons is substantially reduced in confluent cells. Thus, it is conceivable to cure cells of replicons by culturing them at a high density.

Replicon inhibitory agents inhibit replicon activity or select against a cell containing a replicon. An example of such an agent is IFN-α. Other HCV inhibitory compounds may also be employed. HCV inhibitor compounds are described, for example, in Llinas-Brunet, et al., 2000. Bioorg Med Chem. Lett. 10(20), 2267-2270.

The ability of a cured cell to be a replicon enhanced cell can be measured by introducing a replicon into the cell and determining efficiency of subsequent replicon maintenance and activity.

EXAMPLES

Examples are provided below to further illustrate different features of the present invention. The examples also illustrate useful methodology for practicing the invention. These examples do not limit the claimed invention.

Example 1: Techniques

5

15

20

35

This example illustrates the techniques employed for producing and analyzing adaptive mutations and replicon enhanced cells.

Manipulation of Nucleic Acids and Construction of Recombinant Plasmids

Manipulation of nucleic acids was done according to standard

protocols. (Sambrook, et al., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed.

Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) Plasmid DNA was

prepared from ON culture in LB broth using Qiagen 500 columns according to manufacturer instructions.

5

10

15

20

25

30

Plasmids containing desired mutations were constructed by restriction digestion using restriction sites flanking the mutations or by PCR amplification of the area of interest, using synthetic oligonucleotides with the appropriate sequence. Site directed mutagenesis was carried out by inserting the mutations in the PCR primers. PCR amplification was performed using high fidelity thermostable polymerases or mixtures of polymerases containing a proofreading enzyme. (Barnes, et al., 1994. Proc. Natl. Acad. Sci. 91, 2216-2220.) All plasmids were verified by restriction mapping and sequencing.

pHCVneo17.wt contains the cDNA for an HCV bicistronic replicon identical to replicon I₃₇₇neo/NS3-3'/wt described by Bartenschlager (SEQ. ID. NO. 3) (Lohmann, *et al.*, 1999. *Science 285*,110-113, EMBL-genbank No. AJ242652). The plasmid comprises the following elements: 5' untranslated region of HCV comprising the HCV-IRES and part of the core (nt1-377); neomycin phosphotransferase coding sequence; and EMCV IRES; HCV coding sequences from NS3 to NS5B; 3' UTR of HCV.

Plasmid pHCVNeo17.GAA is identical to pHCVNeo.17, except that the GAC triplets (nt. 6934-6939 of pHCVNeo17 sequence) coding for the catalytic aspartates of the NS5B polymerase (amino acids 2737 and 2738 of HCV polyprotein) were changed into GCG, coding for alanine.

Plasmid pHCVNeo17.m0 is identical to pHCVNeo17, except that the triplet AGC (nt. 5335-5337 of pHCVNeo17 sequence) coding for the serine of NS5A protein (amino acid 2204 of HCV polyprotein) was changed into AGA, coding for arginine.

Plasmid pHCVNeo17.m1 is identical to pHCVNeo17, except that the triplet AAC (nt. 4846-4848 of pHCVNeo17 sequence) coding for the asparagine of NS5A protein (amino acid 2041 of HCV polyprotein) was changed into ACC, coding for threonine.

Plasmid pHCVNeo17.m2 is identical to pHCVNeo17, except that the triplet TCC (nt. 5242-5244 of pHCVNeo17 sequence) coding for the serine of NS5A protein (amino acid 2173 of HCV polyprotein) was changed into TTC, coding for phenylalanine.

Plasmid pHCVNeo17.m3 is identical to pHCVNeo17, except that the triplet TCC (nt. 5314-5316 of pHCVNeo17 sequence) coding for the serine of NS5A protein (amino acid 2197 of HCV polyprotein) was changed into TTC, coding for phenylalanine.

Plasmid pHCVNeo17.m4 is identical to pHCVNeo17, except that the triplet TTG (nt. 5317-5319 of pHCVNeo17 sequence) coding for the leucine of NS5A protein (amino acid 2198 of HCV polyprotein) was changed into TCG, coding for serine.

5

10

15

20

25

30

Plasmid pHCVNeo17.m5 is identical to pHCVNeo17, except that an extra triplet AAA coding for lysine was inserted after the triplet GTG (nt. 4840-4843 of pHCVNeo17 sequence), coding for valine 2039 of HCV polyprotein.

Plasmid pHCVNeo17.m6 is identical to pHCVNeo17, except that the triplets GAA and GCC (nt. 2329-2331 and 2764-2766 of pHCVNeo17 sequence) coding for the glutamic acid and the alanine of NS3 protein (amino acid 1202 and 1347 of HCV polyprotein) were changed respectively into GGA and ACC, coding for glycine and threonine. The triplet TCC (nt. 5242-5244 of pHCVNeo17 sequence) coding for the serine of NS5A protein (amino acid 2173 of HCV polyprotein) was changed into TTC, coding for phenylalanine; an extra adenosine was inserted into the EMCV IRES (after the thymidine 1736 of the replicon sequence).

Plasmid pHCVNeo17.m7 is identical to pHCVNeo17, except that the triplet AAC (nt. 4846-4848 of pHCVNeo17 sequence) coding for the asparagine of NS5A protein (amino acid 2041 of HCV polyprotein) was changed into ACC, coding for threonine; the triplet TCC (nt. 5242-5244 of pHCVNeo17 sequence) coding for the serine of NS5A protein (amino acid 2173 of HCV polyprotein) was changed into TTC, coding for phenylalanine.

Plasmid pHCVNeo17.m8 is identical to pHCVNeo17, except that the triplet AAC (nt. 4846-4848 of pHCVNeo17 sequence) coding for the asparagine of NS5A protein (amino acid 2041 of HCV polyprotein) was changed into ACC, coding for threonine; the triplet TCC (nt. 5314-5316 of pHCVNeo17 sequence) coding for the serine of NS5A protein (amino acid 2197 of HCV polyprotein) was changed into TTC, coding for phenylalanine.

Plasmid pHCVNeo17.m9 is identical to pHCVNeo17, except that the triplet AAC (nt. 4846-4848 of pHCVNeo17 sequence) coding for the asparagine of NS5A protein (amino acid 2041 of HCV polyprotein) was changed into ACC, coding

PCT/EP02/00526 WO 02/059321

for threonine; the triplet TTG (nt. 5317-5319 of pHCVNeo17 sequence) coding for the leucine of NS5A protein (amino acid 2198 of HCV polyprotein) was changed into TCG, coding for serine.

Plasmid pHCVNeo17.m10 is identical to pHCVNeo17, except that the triplet GAA (nt. 2329-2331 of pHCVNeo17 sequence) coding for the glutamic acid of NS3 protein (amino acid 1202 of HCV polyprotein) was changed into GGA, coding for glycine; an extra triplet AAA coding for lysine was inserted after the triplet GTG (nt. 4840-4843 of pHCVNeo17 sequence), coding for valine 2039 of HCV polyprotein.

10 Plasmid pHCVNeo17.m11 is identical to pHCVNeo17, except that the triplet TCC (nt. 5314-5316 of pHCVNeo17 sequence) coding for the serine of NS5A protein (amino acid 2197 of HCV polyprotein) was changed into TTC, coding for phenylalanine. The triplet GCC (nt. 5320-5322 of pHCVNeo17 sequence) coding for the alanine of NS5A protein (amino acid 2199 of HCV polyprotein) was changed into ACC coding for threonine.

15

20

25

30

Plasmid pHCVNeo17.m12 is identical to pHCVNeo17, except that the triplet AAC (nt. 4846-4848 of pHCVNeo17 sequence) coding for the asparagine of NS5A protein (amino acid 2041 of HCV polyprotein) was changed into ACC, coding for threonine; the triplet TCC (nt. 5314-5316 of pHCVNeo17 sequence) coding for the serine of NS5A protein (amino acid 2197 of HCV polyprotein) was changed into TTC, coding for phenylalanine. The triplet GCC (nt. 5320-5322 of pHCVNeo17 sequence) coding for the alanine of NS5A protein (amino acid 2199 of HCV polyprotein) was changed into ACC coding for threonine.

Plasmid pHCVNeo17.m13 has the same mutations as pHCVNeo17.m8, but also an extra adenosine inserted into the EMCV IRES (after the thymidine 1736 of the replicon sequence).

Plasmid pHCVNeo17.m14 has the same mutations as pHCVNeo17.m11, but also an extra adenosine inserted into the EMCV IRES (after the thymidine 1736 of the replicon sequence).

Plasmid pHCVNeo17.m15 is identical to pHCVNeo17, except that the triplet GCC (nt. 5320-5322 of pHCVNeo17 sequence) coding for the alanine of NS5A protein (amino acid 2199 of HCV polyprotein) was changed into ACC coding for threonine.

Plasmid pRBSEAP.5 is a pHCVNeo.17 derivative where the Neo coding sequence has been replaced with the sequence coding for the human placental alkaline phosphatase corresponding to nucleotides 90-1580 of pBC12/RSV/SEAP plasmid. (Berger, et al., 1988. Gene 66, 1-10.)

5

10

15

RNA Transfection

Transfection was performed using Huh-7 cells. The cells were grown in Dulbecco's modified minimal essential medium (DMEM, Gibco, BRL) supplemented with 10% FCS. For routine work, cells were passed 1 to 5 twice a week using 1x trypsin/EDTA (Gibco, BRL).

Plasmids were digested with the ScaI endonuclease (New England Biolabs) and transcribed *in vitro* with the T7 Megascript kit (Ambion). Transcription mixtures were treated with DNase I (0.1 U/ml) for 30 minutes at 37°C to completely remove template DNA, extracted according to the procedure of Chomczynski (Chomczynski, *et al.*, 1987. *Anal. Biochem. 162*, 156-159), and resuspended with RNase-free phosphate buffered saline (rfPBS, Sambrook, *et al.*, 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

RNA transfection was performed as described by Liljestrom, et al., 20 1991. J. Virol. 6, 4107-4113, with minor modifications. Subconfluent, actively growing cells were detached from the tissue culture container using trypsin/EDTA. Trypsin was neutralised by addition of 3 to 10 volumes of DMEM/10%FCS and cells were centrifuged for 5 minutes at 1200 rpm in a Haereus table top centrifuge at 4°C. Cells were resuspended with ice cold rfPBS by gentle pipetting, counted with a 25 haemocitometer, and centrifuged as above. rfPBS wash was repeated once and cells were resuspended at a concentration of 1-2 x 10⁷ cell/ml in rfPBS. Aliquots of cell suspension were mixed with RNA in sterile eppendorf tubes. The RNA/cell mixture was immediately transferred into the electroporation cuvette (precooled on ice) and pulsed twice with a gene pulser apparatus equipped with pulse controller (Biorad). 30 Depending on the experiment, 0.1, 0.2 or 0.4 cm electrode gap cuvettes were used, and settings adjusted (Table 3).

TABLE 3

Cuvette	Volume	Voltage	Capacitance	Resistance	RNA
gap (cm)	(μl)	(Volts)	(μFa)	(ohm)	(µg)
0.1	70	200	25	infinite	1-10
0.2	200	400	25	infinite	5-20
0.4	800	800	25	infinite	15-100

After the electric shock, cells were left at room temperature for 1-10 minutes (essentially the time required to electroporate all samples) and subsequently diluted with at least 20 volumes of DMEM/10%FCS and plated as required for the experiment. Survival and transfection efficiency were monitored by measuring the neutral red uptake of cell cultured for various days in the absence or in the presence of neomycin sulfate (G418). With these parameters, survival of Huh-7 cells was usually 40-60% and transfection efficiency ranged between 40% and 100%.

Sequence Analysis of Replicon RNAs

15

20

25

The entire NS region was recloned from 3 different transfection experiments performed with HCVNeo.17 RNA. RNA was extracted from selected clones either using the Qiagen RNAeasy minikit following manufacturer instructions or as described by Chomczynski, et al., 1987. Anal. Biochem. 162, 156-159.

Replicon RNAs (5 µg of total cellular RNA) were retro-transcribed using oligonucleotide HCVG34 (5'- ACATGATCTGCAGAGAGGCCAGT-3'; SEQ. ID. No. 4) and the Superscript II reverse transcriptase (Gibco, BRL) according to manufacturer instructions, and subsequently digested with 2 U/ml Ribonuclease H (Gibco BRL). The cDNA regions spanning from the EMCV IRES to the HCV 3' end were amplified by PCR using oligonucleotides HCVG39 (5'-GACASGCTGTGATAWATGTCTCCCCC-3'; SEQ. ID. NO. 5) and CITE3 (5'-TGGCTCTCCTCAAGCGTATTC -3'; SEQ. ID. NO. 6) and the LA Taq DNA polymerase (Takara LA Taq).

Amplified cDNAs were digested with the KpnI endonuclease (New England Biolabs) and the 5.8 kb fragments were gel purified and ligated to the 5.6 kb vector fragment (purified from plasmid pRBSEAP.5 digested with KpnI) using T4

DNA ligase (New England Biolabs) according to manufacturer instructions. Ligated DNAs were transformed by electroporation in DH10B or JM119 strains of *E. coli*.

In the case of NS5A region, total RNA isolated from 3 clones, (HB77, HB60 and HB68) was extracted and used for RT-PCR. 5μ g of total RNA plus 20 pmole of AS61 oligo (5'-ACTCTCTGCAGTCAAGCGGCTCA-3', RT antisense oligo; SEQ. ID. NO. 7) were heated 5 minutes at 95°C, then DMSO (5% f.c.), DTT (10 mM f.c.), 1 mM dNTP (1 mM f.c.), 1x Superscript buffer (1 x f.c.), and 10 u Superscript (Gibco) were added to a total volume of 20 μ l and incubated 3 hours at 42°C. 2μ l of this RT reaction were used to perform PCR with oligos S39 (5'-

CAGTGGATGAACCGGCTGATA-3', sense; SEQ. ID. NO. 8) or S41 (5'-GGGGCGACGGCATCATGCAAACC-3', sense; SEQ. ID. NO. 9) and B43 (5'-CAGGACCTGCAGTCTGTCAAAGG-3', antisense; SEQ. ID. NO. 10) using Elongase Enzyme Mix (Gibco) according the instruction provided by the manufacturer. The resulting PCR fragment was cloned in pCR2.1 vector using the TA Cloning kit (Invitrogen) and transformed in Top10F' bacterial strain.

Plasmid DNA was prepared from ON culture of the resulting ampicillin resistant colonies using Qiagen 500 columns according to manufacturer instructions. The presence of the desired DNA insert was ascertained by restriction digestion, and the nucleotide sequence of each plasmid was determined by automated sequencing. Nucleotide sequences and deduced amino acids sequences were aligned using the GCG software.

TaqMan

5

10

15

20

TaqMan analysis was typically performed using 10 ng of RNA in a

reaction mix (TaqMan Gold RT-PCR kit, Perkin Elmer Biosystems) either with HCV

specific oligos/probe (oligo 1: 5'-CGGGAGAGCCATAGTGG-3'; SEQ. ID. NO. 11,

oligo 2: 5'-AGTACCACAAGGCCTTTCG-3'; SEQ. ID. NO. 12, probe: 5'
CTGCGGAACCGGTGAGTACAC-3'; SEQ. ID. NO. 13) or with human GAPDH

specific oligos/probe (Pre-Developed TaqMan Assay Reagents, Endogenous Control

Human GAPDH, Part Number 4310884E, Perkin Elmer Biosystems). PCR was

performed using a Perkin Elmer ABI PRISM 7700 under the following conditions: 30

minutes at 48°C (the RT step), 10 minutes at 95°C and 40 cycles: 15 seconds at 95°C

and 1 minute at 60°C. Quantitative calculations were obtained using the Comparative

C_T Method (described in User Bulletin #2, ABI PRISM 7700 Sequence Detection

System, Applied Biosystem, Dec 1997) considering the level of GAPDH mRNA

constant. All calculations of HCV RNA are expressed as fold difference over a specific control.

Antibodies and Immunological Techniques

5

10

15

20

25

30

35

Mouse monoclonal antibody (anti-NS3 mab10E5/24) were produced by standard techniques. (Galfré and Milstein, 1981. *Methods in Enzymology 73*, 1-46.) Purified recombinant protein was used as an immunogen. (Gallinari, *et al.*, 1999. *Biochemistry 38*, 5620-5632.)

For Cell-ELISA analysis, transfected cells were monitored for expression of the NS3 protein by ELISA with the anti-NS3 mab 10E5/24. Cells were seeded into 96 well plates at densities of 40,000, 30,000, 15,000 and 10,000 cells per well and fixed with ice-cold isopropanol at 1, 2, 3 and 4 days post-transfection, respectively. The cells were washed twice with PBS, blocked with 5% non-fat dry milk in PBS + 0.1% Triton X100 + 0.02% SDS (PBSTS) and then incubated overnight at 4°C with 10E5/24 mab diluted 1:2000 in Milk/PBSTS. After washing 5 times with PBSTS, the cells were incubated for 3 hours at room temperature with anti-mouse IgG Fc specific alkaline phosphatase conjugated secondary antibody (Sigma A-7434), diluted 1:2000 in Milk/PBSTS. After washing again as above, the reaction was developed with *p*-nitrophenyl phosphate disodium substrate (Sigma 104-105) and the absorbance at 405 nm read at intervals.

The results were normalized by staining with sulforhodamine B (SRB Sigma S 1402) to determine cell numbers. The alkaline phosphatase substrate was removed from the wells and the cells washed with PBS. The plates were then incubated with 0.4% SRB in 1% acetic acid for 30 minutes (200 µl/well), rinsed 4 times in 1% acetic acid, blotted dry and then 200 µl/well of 10mM Tris pH 10.5 added. After mixing, the absorbance at 570 nm was read.

Neutral Red/ Crystal Violet Staining of Foci

The survival of transfected cells in the absence or presence of G418 was monitored by staining of foci/clones with neutral red *in vivo* with subsequent crystal violet staining. The medium was removed from the cells and replaced with fresh medium containing 0.0025% neutral red (Sigma N2889) and the cells incubated for 3 hours at 37°C. Cells were washed twice with PBS, fixed in 3.5% formaldehyde for 15 minutes, washed twice again in PBS and then with distilled water and the number of (live) foci counted. The cells could then be re-stained with crystal violet

by incubating with an 0.1% crystal violet (Sigma C0775) solution in 20% methanol for 20 minutes at room temperature, followed by 3 washes in 20% methanol and a wash with distilled water.

5 Preparation Of Cells Cured Of Endogenous Replicon

Replicon enhanced cells designated 10IFN and Cl.60/cu were produced using different HCV inhibitory agents. Based on the techniques described herein additional replicon enhanced clones can readily be obtained.

10 IFN-α2b. Huh-7 cells containing HCV replicons (designated HBI10, HBIII4, HBIII27 and HBIII18) were cultured for 11 days in the presence of 100 U/ml recombinant human IFN-α2b (Intron-A, Schering-Plough), and subsequently for 4 days in the absence of IFN-α2b. At several time points during this period, the clones were analyzed for the presence of HCV proteins and RNA by Western and Northern

15 blotting. After 7 days of incubation with IFN-α2b, HCV proteins could no longer be detected in any of these clones by Western blotting and similar effects were seen with RNA signals in Northern blots. IFN-α2b treated cells were stored in liquid nitrogen until used for transfection experiments.

Cl.60/cu was obtained by curing a Huh-7 cell of a replicon using an HCV inhibitory compound. The presence of HCV RNA was determined using PCR (TaqMan) at 4, 9, 12 and 15 days. From day 9 the amount of HCV RNA was below the limit of detection. To further test the disappearance of the replicon, 4 million cells of cured Clone 60 cells (after the 15 days of treatment) were plated in the presence of 1 mg/ml G-418. No viable cells were observed, confirming that absence of HCV replicons able to confer G-418 resistance.

<u>Example 2: Selection and Characterization of Cell Clones Containing Functional HCV Replicons</u>

Huh-7 cells (2-8x10⁶) were transfected by electroporation with *in vitro* transcribed replicon RNAs (10–20 μg), plated at a density ranging from 2.5x10³ to 10x10³/cm², and cultured in the presence of 0.8-1 mg/ml G418. The majority of replicon transfected cells became transiently resistant to G418 and duplicated normally for 7 to 12 days in the presence of the drug, while cells transfected with irrelevant RNAs and mock transfected cells did not survive more than 7 days (data not shown). Transient resistance to G418 was likely due to persistence of the Neo protein

expressed from the transfected RNA, since it was observed also with mutated replicons unable to replicate. Approximately 2 weeks after transfection, transient resistance declined, most cells died and small colonies of cells permanently resistant to the antibiotic became visible in samples transfected with HCVNeo.17 RNA, but not in cells transfected with other replicon RNAs.

In several experiments, the frequency of G418 resistant clones ranged between 10 and 100 clones per 10⁶ transfected cells. About 20 G418 resistant colonies were isolated, expanded and molecularly characterized. PCR and RT-PCR analysis of nucleic acids indicated that all clones contained HCV RNA but not HCV DNA, demonstrating that G418 resistance was due to the presence of functional replicons (data not shown). This result was confirmed by Northern blot analysis and metabolic labeling with 3H-uridine, which revealed the presence of both genomic and antigenomic HCV RNAs of the expected size (data not shown). Lastly, western blot, immunoprecipitation and immunofluorescence experiments showed that these clones expressed all HCV non-structural proteins as well as Neo protein (data not shown).

Clones differed in terms of cell morphology and growth rate. Replicon RNA copy number (500-10000 molecules/cell) and viral protein expression also varied between different clones (data not shown). However, the amount of replicon RNA and proteins also varied with passages and with culture conditions and was higher when cells were not allowed to reach confluency, suggesting that replicons replicated more efficiently in dividing cells than in resting cells. Processing of the viral polyprotein occurred with kinetics similar to those observed in transfected cells.

Interestingly, in all tested clones HCV replication was efficiently inhibited by treating the cells with IFN- α 2b. The EC50 was between 1 and 10 U/ml, depending on the clone.

Example 3: Identification of Adaptive Mutations

5

10

15

20

25

30

The low number of G418 resistant clones derived from HCVNeo.17 RNA transfection suggested that replication could require mutation(s) capable of adapting the replicon to the host cell (adaptive mutations) and/or that only a small percentage of Huh-7 cells were competent for HCV replication. To verify the first hypothesis, mutations in replicons RNAs derived from selected cell clones were identified.

RNA sequences for different replicons were determined using standard techniques. Such techniques involved isolating RNA from several independent clones, reverse transcription to produce cDNA, amplifying cDNAs by PCR and cloning into an appropriate vector. The cDNA spanning almost the entire HCV NS region (126 bp at the 3' end of the EMCV IRES and 5650 bp of the HCV NS region (*i.e.*, the entire NS ORF and 298 nucleotides at the 3' end) from 5 clones (HBI10, HBIII12, HBIII18, HBIII27, HBIV1) were recloned and sequenced. In addition, the NS5A coding region (nt. 4784-6162) from 3 additional clones (HB 77, HB 68 and HB 60) were recloned and sequenced.

To discriminate mutations present in the replicon RNA from those derived from the cloning procedure, at least 2 isolates derived from independent RT-PCR experiments were sequenced for each cell clone. Comparison of the nucleotide sequences with the parental sequence indicated that each isolate contained several mutations (Tables 4A and 4B).

15

10

TABLE 4A

Cell clone	НВІ	II 12	НВІ	II 18	НВ	I 10	НВІ	II 27
isolate	4	29	28	61	12	43	13	72
	1674-	1674-	1674-	1674-	1674-	1674-	1674-7460	1674-7460
	7460	7460	7460	7460	7460	7460		
EMCV IRES 126 bp	A @ 1736	A @ 1736		C 1752 T				T 1678 C
•	5 5000 5							
NS3							G 2009 C	G 2009 C
1895 bp	A 2698 G	C 2505 T	C 2196 A	A 2338 G	A 2330 G	A 2882 G	T 2015 C	C 2052 A
	G 2764 A	G 2764 A	T 3023 A	C 2616 T	T 2666 C	T 3673 C	C 2336 G	G 2644 A
	A 3256 G	T 3085 C	T 3134 C	A 2664 G	T 3395 C		A 3130 T	C 2803 A
	T 3273 C		C 3267 T	A 3148 G			A 3401 G	T 2823 A
				T 3286 C			A 3518 C	T 3692 C
				C 3615 T				
				C 3657 T				,
NS4A	T 3790 C		A 3847 G	T 3827 A	T 3742 C		A 3743 G	A 3797 G
161 bp					1			
NS4B	T 3869 C	C 4283 T	G 4300 A	A 4136 G	T 4290 C	A 4053 G	G 3880 A	C 4547 T
782 bp	A 4107 G	C 4429 T		A 4261 G		A 2496 C	T 4200 C	
•	T 4185 C			G 4309 A		T 4316 G	A 4366 G	
	Λ 4428 G			A 4449 G				

TABLE 4A

Cell clone	НВІ	II 12	нві	II 18	НВ	I 10	HBI	II 27
isolate	4	29	28	61	12	43	13	72
	1674- 7460	1674- 7460	1674- 7460	1674- 7460	1674- 7460	1674- 7460	1674-7460	1674-7460
NS5A	A 4847 C	G 4728 A	C 5243 T	C 4729 A	A 4694 T	A 4675 G	A 4855 G	A 4888 G
1340 bp	G 5158 A	A 4845 G	A 5486 G	T 4993 C	AAA @ 4842	A 4761 G	C 5006 T	C 4985 T
	G 5175 C	C 5243 T	C 5596 T	G 5095 A	T 5237 C	AAA @ 4842	T 5318 C	T 5030 A
	C 5243 T	G 5512 T	G 5823 A	T 5334 C		T 5368 C	A 5574 G	T 5090 A
	C 5390 T	A 5521 G		A 5374 T			G 5866 A	T 5318 C
1 .	A 5719 G	A 5600 G		T 5379 A				A 5328 G
		A 5740 C		T 5480 C				A 5399 G
				A 5513 G				A 5574 G
				T 5977 C				
NS5B	T 6316 C	A 6406 G	T 6074 C	A 6150 G	A 6911 G	A 5986 G	G 6479 C	G 6156 A
1477 bp	T 6589 C	G 6756 A	A 6541 G	A 6218 G		T 6099 C	C 6870 T	G 7434 A
	T 7370 C	G 6963 T	A 6732 G	T 7352 A		C 6141 T	A 7213 G	T 7444 C
			A 7350 T			G 6463 A	T 7448 C	
			A 7359 G			C 6849 T	•	
						T 6865 C		

Clone name and isolate number are indicated in the first and second row, respectively.

The first and the last nucleotide of the region that was recloned and sequenced are indicated in the third row.

Nucleotide (IUB code) substitutions are indicated with the original nucleotide, its position and mutated nucleotide.

Nucleotide(s) insertions are indicated with the nucleotide(s), the symbol @ and the position of the nucleotide preceding insertion.

Numbering refers to the first nucleotide of the replicon sequence (EMBL-genbank No. AJ242652). The region in which mutations are located and the nucleotide length of each region are indicated in the left most column.

Silent mutations are in italic.

Non sense mutations are underlined.

15 Consensus mutations are bold.

5

TABLE 4B

Cell clone	НВ	IVI	HE	3 77	НЕ	3 68	HE	60
isolate	85	93	10	14	42	I	13	7
	1674- 7460	1674- 7460	4784- 6162	4465- 6162	4784- 6162	4465- 6162	4784-6162	4784-6162
EMCV IRES 126 bp		A @ 1736						
NS3 1895 bp	A 3403 G	A 2572 G A 3454 G						

TABLE 4B

Cell clone	НВ	IVI	HE	3 77	HE	8 68	HB	60
isolate	85	93	10	14	42	1	13	7
	1674- 7460	1674- 7460	4784- 6162	4465- 6162	4784- 6162	4465- 6162	4784-6162	4784-6162
NS4A								
161 bp						į .		
NS4B	A 4084 G	C 3892 T						
782 bp								į
NS5A	T 4742 C	A 4847 C	C 4813 T	A 4699 C	T 5171 G	T 4587 C	A 4821 G	C 5337 G
1340 bp	C 5315 T	A 5225 G	G 5060 C	A 5161 G	C 5298 T	T 4972 C	G 5320 A	C 5551 T
	<u>G 5431 T</u>	C 5315 T	C 5337 A	C 5337 A	C 5337 A	A 5094 G	A 5414 G	G 5806 A
		G 5320 A					T 5601 G	
1	T 5797 C	T 5356 A		T 5977 C	A 5969 G	G 5320 A	C 5808 T	
		G 5523 A				C 5532 T	İ	
		T 5888 A			~-			
NS5B	T 6144 A							
1477 bp	A 6365 G						j	
	A 6656 G	T 7171 C						
	A 6677 G							
	T 6855 C							
	T 6947 A							i
	T 6997 C							
	G 7041 T		j	ĺ				
	A 7187 C							

See Table 4A legend.

5

10

15

The frequency of mutations ranged between 1.7×10^{-3} and 4.5×10^{-3} (average 3×10^{-3}). The majority of mutations were nucleotide substitutions, although insertions of 1 or more nucleotides were also observed (Tables 4A and 4B).

Approximately 85% of the mutations found only in 1 isolate (non-consensus) were randomly distributed in the recloned fragment, and possibly include mis-incorporation during the PCR amplifications. Conversely, the remaining 15% of the mutations were common to 2 or more isolates derived from independent RT-PCR experiments (consensus mutations), and presumably reflected mutations present in the template RNA.

Consensus mutations were found in all isolates and were either common to isolates derived from the same clone (consensus A), or to isolates derived from different clones (consensus B). Analysis of additional isolates derived from the same cell clones indicated that consensus A mutations were not always present in all isolates derived from one clone (data not shown). This observation, together with the

presence of consensus B mutations, suggests that, even within a single cell clone, replicons exist as quasi-species of molecules with different sequences.

At variance with non-consensus mutations, consensus mutations were not randomly distributed but were clustered in the regions coding for the NS5A protein (frequency 1×10^{-3}) and for the NS3 protein (frequency 0.5×10^{-3}). Only one consensus mutation was found in the region coding for the NS5B protein (frequency 0.1×10^{-3} nucleotides) and none in the regions coding for NS4A and NS4B. Interestingly, I consensus mutation was observed also in the EMCV IRES.

With the exception of 2 silent mutations found in NS5A and NS5B, consensus mutations occurring in the NS region resulted in changes in the deduced amino acid sequence (Tables 5A and 5B). Noticeably, these amino acid changes occurred in residues that are conserved in all or most natural HCV isolates. Interestingly, clones HB 77 and HB 60 displayed different nucleotide substitutions (C5337A and C5337G, respectively) resulting in the same amino acidic mutation (S 2204 R).

TABLE 5A

Cell clone	НВІ	II 12	HBI	11 18	НВ	I 10	НВ	III 27
isolate	4	29	28	61	12	43	13	72
NS3	G 1095 A A 1347 T	E 1202 G A 1347 T			E 1202 G	E 1202 G	G 1095 A	G 1095 A
NS4A								-
NS4B								
	N 2041 T S 2173 F	S 2173 F	S 2173 F	E 2263	K @ 2039		L 2198 S R 2283 R	L 2198 S R 2283 R
NS5B								

See Table 4A legend.

20

5

TABLE 5B

Cell clone	НВ	IV1	HB	77	HE	3 68	Н	B 60
isolate	85	93	10	14	42	1	13	7
NS3								
NS4A								
NS4B								
NS5A	S 2197 F	N 2041 T S 2197 F A 2199 T	S 2204 R	S 2204 R	S 2204 R	A 2199 T	A 2199 T	S 2204 R
NS5B	N 2710 N	N 2710 N	-					<u> </u>

See Table 4A legend.

10

5 Example 4: Functional Characterization of Consensus Mutations

The identification of consensus mutations in recloned replicons indicated that replication proficiency of replicon RNAs contained in selected cell clones depended from the presence of such mutations. To substantiate this hypothesis, the effect of several consensus mutations on replication were analyzed.

Consensus mutations found in the NS5A region were more closely analyzed. Consensus mutations were segregated from the non-consensus ones, and pHCVNeo.17 derivatives containing single or multiple consensus mutations were constructed (Table 6).

TABLE 6

Construct		Consensus mutations		G418 cfw/10 ⁵ transfected cells
	NS3	NS5A	EMCV IRES	
pHCVNeo17.wt				0-3
pHCVNeo17.GAA				0
pHCVNeo17.m0		S2204R		30-130
pHCVNeo17.ml		N2041T		0-3
pHCVNeo17.m2		S2173F		15-60
pHCVNeo17.m3		S2197F		160-500
pHCVNeo17.m4		L2198S		30-50

TABLE 6

Construct		Consensus mutations		G418 cfu/10 ⁵
Ì				transfected
				cells
}	NS3	NS5A	EMCV IRES	
pHCVNeo17.m5	i	<u>K@2039</u>		25-55
pHCVNeo17.m6	E1202G; A1347T	S2173F	Extra A	13-100
pHCVNeo17.m7		N2041T; S2173F		0-1
pHCVNeo17.m8		N2041T; S2197F		360-500
pHCVNeo17.m9		N2041T; L2198S		140-170
pHCVNeo17.m10	E1202G	<u>K@2039</u>		1060
pHCVNeo17.m11		S2197F; A2199T		900
pHCVNeo17.m12		N2041T; S2197F; A2199T		>1000
pHCVNeo17.m13		N2041T; S2197F	Extra A	100
pHCVNeoi7.m14		S2197F; A2199T	Extra A	>500
pHCVNeo17.m15		A2199T		300-600

Huh-7 cells (2x10⁶) were transfected with 10 µg of RNA transcribed from the indicated constructs. Approximately 2x10⁵ cells were plated in a 10 cm tissue culture dish and cultured with 1 mg/ml G418 for 20 days.

Colonies surviving selection were stained with crystal violet and counted.

5

10

RNAs transcribed *in vitro* from these constructs were transfected in Huh-7 cells and the affect on replication was estimated by counting neomycin resistant colonies (G418 cfu). As shown in Table 6, all but 1 construct containing single consensus mutations showed a significant increase on G418 cfu efficiency, thus indicating that the corresponding mutations improved replication. Noticeably, 2 mutants containing single mutations in NS5A (m3 and m15) were clearly more effective than all other single mutants. Results of mutants containing 2 or more mutations, indicated the presence of a synergistic effect in some combinations (m8, m9, m11 and possibly m10), but also a slightly antagonistic effect in 1 mutant (m7).

15

20

Example 5: Replicon Replication in the Absence of Selection

Replication of HCV replicons in the absence of a G418 selection was detected using quantitative PCR (TaqMan). At 24 hours post-transfection a large amount of replicon RNA was detected in cells transfected with all replicons, including the GAA control replicon containing mutations in the catalytic GDD motif of the NS5B polymerase. This result suggested that analysis at very early time points (up to 48 hour post-transfection) essentially measured the input RNA. Northern blot analysis also indicated that after 24 hours the majority of the transfected RNA was degraded intracellularly (data not shown).

Analysis at later time points showed that the amount of replicon RNA was considerably reduced at 4 days and eventually became undetectable (6/8 days) in cells transfected with replicon HCVNeo17.wt, but was still high in cells transfected with replicons m0, m3 and m15 (Table 7). At day six, that the amount of replicon RNA became undetectable in cells transfected with replicon HCVNeo17.wt, m0, and m2, but was detectable in cells transfected with replicon m3 and m15 (Table 7).

TABLE 7

Name	Hu H7				
	RNA equ.	RNA equ.			
	day 4	day 6			
Wt	1 x	1 x			
hcvneo17.m0	3 x	1 x			
hcvneo17.m2	1 x	1 x			
hcvneo17.m3	5 x	3 x			
hcvneo17.m15	6 x	5 x			

10

15

20

5

Persistence of m0, m3 and m15 replicons RNA was abolished by treatment with interferon-α or with an HCV inhibitory compound (data not shown). Moreover, RNA persistence was not observed with mutated replicons carrying the NS5B GAA mutation besides adaptive mutations (data not shown). Taken together, these results demonstrated that quantitative PCR could be used to monitor replication at early times post-transfection, and can be used to evaluate the replication proficiency of replicon RNAs containing mutations.

Comparison of the results shown in Tables 6 and 7, indicated that there was a good correlation between the amount of replicon RNA detected by TaqMan and the G418 cfu efficiency. Nonetheless, some mutants (m2, m3) showed a pronounced effect on G418 cfu efficiency, and little if any effect on early replication as measured by TaqMan PCR, while other mutants (m0) showed the reverse behavior.

Example 6: HCV Replicon Enhanced Cells

HCV replicon enhanced cells were produced by introducing an HCV replicon into a host, then curing the host of the replicon. Adaptive mutations (or combinations of them) by themselves increased up to 2 orders of magnitude the G418 cfu efficiency and enhanced early replication comparably. Nonetheless, even with the most effective mutants, only a small percentage of transfected cells (<5 %, data not shown) gave rise to G418 resistant clones containing functional replicons. This observation was attributed, at least in part to a low cloning efficiency of Huh-7 cells (data not shown), and only a fraction of Huh-7 cells being competent for replication.

Several clones were cured of endogenous replicons by treating them for about 2 weeks with IFN- α or with a HCV inhibitory compound. Analysis at the end of the treatment showed that neither viral proteins nor replicon RNA could be detected.

Cured cells (10IFN and Cl.60/cu) were transfected with mutated replicons and replication efficiency was determined by counting neomycin resistant clones (10IFN) or by TaqMan (10IFN and Cl.60/cu). As shown in Table 8, for all tested replicons the G418 cfu efficiency in 10IFN cells was at least 5 fold higher than in parental Huh-7 cells. This increase in G418 cfu efficiency was particularly relevant for a subset of mutants (m3, m5, m8, m9, m15).

20

5

10

15

TABLE 8

Construct	Consensus mutations			G418 cfu/10 ⁵ transfected cells	
	NS3	NS5A	EMCV IRES		
pHCVNeo17.wt				12 - 56	
pHCVNeo17.GAA			ĺ	0 .	
pHCVNeo17.m0		S2204R		180 - 1000	
pHCVNeo17.ml		N2041T		8 - 13	
pHCVNeo17.m2		S2173F		2000	
pHCVNeo17.m3		S2197F		1600 - 3000	
pHCVNeo17.m4		L2198S	i	190 - 650	
pHCVNeo17.m5		K@2039		1600 - 3000	
pHCVNeo17.m6	E1202G; A1347T	S2173F	extra A	600 - 2000	
pHCVNeo17.m7		N2041T; S2173F	i	170 - 800	
pHCVNeo17.m8		N2041T; S2197F		> 4000	
pHCVNeo17.m9		N2041T; L2198S	j	1400 - 3000	
pHCVNco17.m10	E1202G	K@2039	j	>4000	
pHCVNeo17.m11		S2197F; A2199T		>4000	

TABLE 8

Construct	Consensus mutations		G418 cfu/10 ⁵ transfected cells	
	NS3	NS5A	EMCV IRES	
pHCVNeo17.m12		N2041T; S2197F; A2199	T	>4000
pHCVNeo17.m13		N2041T; S2197F	extra A	>4000
pHCVNeo17.m14		S2197F; A2199T	extra A	>4000
pHCVNeo17.m15		A2199T		> 4000

10IFN cells (2x10⁶) were transfected with 10 µg of RNA transcribed from the indicated constructs. Approximately 2x10⁵ cells were plated in a 10 cm tissue culture dish and cultured with 1 mg/ml G418 for 20 days.

Colonies surviving selection were stained with crystal violet and counted.

Strikingly, the best mutants yielded a number of G418 resistant clones ranging between 20 and 80% of the cell clones which grew in the absence of G418 (data not shown), thus indicating that the majority of 10IFN cells were competent for replication. This result was confirmed by TaqMan analysis (Table 9), in which the fold increase versus the parental Huh-7 cells was very high. The data indicates that replicons carrying adaptive mutations replicate vigorously in replicon enhanced cells such as 10IFN and Cl.60/cu.

15

10

5

TABLE 9

Name	10IFN		C1.60/cu.	
	RNA equ.	RNA equ.	RNA equ.	RNA equ.
	Day 4	day 6	day 4	Day 6
Wt	1 x	1 x	1 x	l x
hcvneo17.m0	46 x	12 x	78 x	512 x
hcvneo17.m2	2 x	2 x	1 x	2 x
hcvneo17.m3	68 x	49 x	19 x	392 x
hcvneo17.m15	247 x	80 x	268 x	5518 x

Expression of viral proteins was determined in replicon enhanced cells
using an ELISA assay designed to detect the NS3 protein in transfected cells plated in
96 wells microtiter plates (Cell-ELISA). As shown in Table 10, 24 hours posttransfection cells transfected with all tested replicons expressed low but detectable
levels of the NS3 protein.

TABLE 10

		NS3 arbitrary units			
	24 h	p.t.	961	ı p.t.	
Name	•	+ IFN	-	+IFN	
Construct					
Mock	1	1	1	1	
pHCVNeo17.wt	3.7	4.2	1.2	1.3	
pHCVNeo17.GAA	3.1	3.2	1.1	1	
pHCVNeo17.m0	3.4	3.2	9.9	0.8	
pHCVNeo17.m3	5.7	4.6	4.7	1.5	
pHCVNeo17.m8	6.6	5.1	15.1	1.4	
pHCVNeo17.m10	8	5.6	9.2	1.8	
pHCVNeo17.m11	8.4	6.2	13.6	1.8	

10IFN cells (2x10⁶) were transfected with 10 µg of RNA transcribed from the indicated constructs. Cells were plated in 96 wells microtiter plates as indicated in Example 1.

5 Where indicated (+IFN), IFN-α (100 U/ml) was added to the culture medium 4 hours post-transfection. At the indicated times post-transfection, cells were fixed and analyzed by Cell-ELISA.

The early expression shown in Table 10 is likely due to translation of transfected RNA, since it was comparable in all replicons (including that carrying the GAA mutation) and was not affected by IFN-α. At 4 days post-transfection, NS3 expression persisted or increased in cells transfected with replicons carrying consensus mutations, but could not be detected anymore in cells transfected with wt and GAA replicons. In addition, NS3 expression was almost completely abolished when cells were cultured in the presence of IFN-α.

Taken together, these results indicated that the level of NS3 expression reflected the replication rate. Indeed, NS3 expression level (Table 10) paralleled the RNA level measured by TaqMan (Table 9). The high replication proficiency of 10IFN cells was further confirmed by immunofluorescence experiments which showed that more than 50% of cells transfected with replicons m8 and m11 expressed high level of viral proteins, and that expression was almost completely abolished by IFN-α.

Example 7: Replication of Full Length Constructs

10

15

20

This example illustrates the ability of a full length HCV genome containing adaptive mutations described herein to replicate in a replicon enhanced host cell. The full length sequence of the HCV isolate Con-1 (EMBL-Genbank No. AJ238799) (plasmid pHCVRBFL.wt) and 2 derivatives containing either the N2041T

and S2173 F mutations (plasmid pHCVRBFL.m8) or the S2197F and A2199T mutations (plasmid pHCVRBFL.m11) were used as starting constructs.

RNAs transcribed from the starting constructs were transfected in 10IFN cells and their replication proficiency was assessed by Cell-ELISA, immunofluorescence and TaqMan. Both constructs containing consensus mutations (pHCVRBFL.m8 and pHCVRBFL.m11) replicated, while no sign of replication was observed with the wt. construct (data not shown).

Example 8: Replicons with Reporter Gene

This example illustrates an HCV replicon containing adaptive mutations and a reporter gene. A pHCVNeo17.wt derivative where the Neo coding region was substituted with that coding for human placental secretory alkaline phosphatase (pRBSEAP5.wt) and a derivative also containing the N2041T and S2173F mutations (plasmid pRBSEAP5.m8) were constructed. RNAs transcribed from these plasmids were transfected in 10IFN cells and their replication proficiency was assessed by measuring secretion of alkaline phosphatase. Analysis of the kinetics of secretion suggested that only plasmid pRBSEAP5.m8 was competent for replication (data not shown).

20 Example 9: SEQ. ID. Nos. 1 and 2

SEQ. ID. NOs. 1 and 2 are provided as follows:

SEQ. ID. NO. 1

5

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKT

25 SERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGS
RPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGV
RVLEDGVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGVYHVTNDCS
NASIVYEAADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNASVPTTTIRR
HVDLLVGAAALCSAMYVGDLCGSVFLVAQLFTFSPRRHETVQDCNCSIYPGH

30 VTGHRMAWDMMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLA
YYSMVGNWAKVLIVMLLFAGVDGGTYVTGGTMAKNTLGITSLFSPGSSQKIQ
LVNTNGSWHINRTALNCNDSLNTGFLAALFYVHKFNSSGCPERMASCSPIDAF
AQGWGPITYNESHSSDQRPYCWHYAPRPCGIVPAAQVCGPVYCFTPSPVVVG
TTDRFGVPTYSWGENETDVLLLNNTRPPQGNWFGCTWMNSTGFTKTCGGPP

CNIGGIGNKTLTCPTDCFRKHPEATYTKCGSGPWLTPRCLVHYPYRLWHYPC TVNFTIFKVRMYVGGVEHRLEAACNWTRGERCNLEDRDRSELSPLLLSTTEW QVLPCSFTTLPALSTGLIHLHQNVVDVQYLYGIGSAVVSFAIKWEYVLLLFLLL ADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAWY 5 IKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLILL TLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPE LIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVOM ALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITW GADTAACGDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQOTRGL 10 LGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLA GPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFV PVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAOG YKVLVLNPSVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPITYSTYGKFLA 15 DGGCSGGAYDIIICDECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSV TVPHPNIEEVALSSTGEIPFYGKAIPIETIKGGRHLIFCHSKKKCDELAAKLSGLG LNAVAYYRGLDVSVIPTSGDVIVVATDALMTGFTGDFDSVIDCNTCVTOTVD FSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRMGIYRFVTPGERPSGMFDSSVL CECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLEFWESVFTGLTHID 20 AHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKPTLHG PTPLLYRLGAVQNEVTTTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAA YCLTTGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAEQ **FKQKAIGLLQTATKQAEAAAPVVESKWRTLEAFWAKHMWNFISGIQYLAGLS** TLPGNPAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAG .25 IAGAAVGSIGLGKVLVDILAGYGAGVAGALVAFKVMSGEMPSTEDLVNLLPA ILSPGALVVGVVCAAILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPE SDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSWLRDVWDWICTVLTD FKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQITGHVKNG SMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVT 30 RVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEV TFLVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSL ASSSASQLSAPSLKATCTTRHDSPDADLIEANLLWRQEMGGNITRVESENKVV ILDSFEPLQAEEDEREVSVPAEILRRSRKFPRAMPIWARPDYNPPLLESWKDPD YVPPVVHGCPLPPAKAPPIPPPRRKRTVVLSESTVSSALAELATKTFGSSESSA 35 VDSGTATASPDQPSDDGDAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEE

ASEDVVCCSMSYTWTGALITPCAAEETKLPINALSNSLLRHHNLVYATTSRSA
SLRQKKVTFDRLQVLDDHYRDVLKEMKAKASTVKAKLLSVEEACKLTPPHS
ARSKFGYGAKDVRNLSSKAVNHIRSVWKDLLEDTETPIDTTIMAKNEVFCVQ
PEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYGFQYSPGQR
VEFLVNAWKAKKCPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQA
IRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAAAACRA
AKLQDCTMLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSAPPGDPPKPE
YDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWL
GNIIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQ
RLHGLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRARLLSQGGR
AATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFVAGYSGGDIYHSLSRARP
RWFMWCLLLLSVGVGIYLLPNR

SEQ. ID. NO. 2:

15 gccagccccgattgggggggacactccaccatagatcactcccctgtgaggaactactgtcttcacgcagaaagcgtcta gccatggcgttagtatgagtgtcgtgcagcctccaggacccccctcccgggagagccatagtggtctgcggaaccggtg agtacaccggaattgccaggacgaccgggtcctttcttggatcaacccgctcaatgcctggagatttgggcgtgccccgcg agactgctagccgagtagtgttgggtcgcgaaaggccttgtggtactgcctgatagggtgcttgcgagtgccccgggaggt ctogtagaccgtgcaccatgagcacgaatcctaaacctcaaagaaaaaccaaacgtaacaccaaccgccgccacagga 20 cgtcaagttcccgggcggtggtcagatcgtcggtggagtttacctgttgccgcgcaggggccccaggttgggtgtgcgcgc gactaggaagacttccgagcggtcgcaacctcgtggaaggcgacaacctatccccaaggctcgccagcccgagggtagg gcctgggctcagcccgggtacccctggcccctctatggcaatgagggcttggggtgggcaggatggctcctgtcaccccgtggctctcggcctagttggggccccacggacccccggcgtaggtcgcgcaatttgggtaaggtcatcgataccctcacgtgc ggcttcgccgatctcatggggtacattccgctcgtcggcgcccccctagggggcgctgccagggccctggcgcatggcgt 25 gtttgaccateccagetteegettatgaagtgegeaacgtateeggagtgtaccatgteacgaacgaetgeteeaacgeaag tgggtagegeteacteceaegetegeggeeaggaaegetagegteeeeactaegaegataegaegeeatgtegatttgete gttggggeggetgetetetgeteegetatgtaegtgggagatetetgeggatetgtttleetegtegeeeagetgtteaeettete 30 gecteg ceggeacg aga cag ta cag gact gea at tget caat at a tecegge cae g tgacag geac acg ta cag ta cae gac a gac a cae gac atatgatgatgaactggtcacctacagcagccctagtggtatcgcagttactccggatcccacaagctgtcgtggatatggtgg egggggeceattggggagteetagegggeettgeetactatteeatggtggggaactgggetaaggttetgattgtgatgeta ctctttgccggcgttgacgggggaacctatgtgacaggggggacgatggccaaaaacaccctcgggattacgtccctctttt caccegggtcatcccagaaaatccagcttgtaaacaccaacggcagctggcacatcaacaggactgccctgaactgcaat 35 gactocotoaacactgggttocttgctgcgctgttotacgtgcacaagttcaactcatctggatgcccagagcgcatggccag

ctgcagccccatcgacgcgttcgctcaggggtgggggcccatcacttacaatgagtcacacagctcggaccagaggcctta aacaacacgcggccgccgcaaggcaactggtttggctgtacatggatgaatagcactgggttcaccaagacgtgcggggg cccccgtgtaacatcggggggatcggcaataaaaccttgacctgccccacggatgcttccggaagcaccccgaggcca cttacaccaagtgtggttcggggccttggttgacacccagatgcttggtcactacccatacaggctttggcactacccctgc gaggagagcgttgtaacctggaggacaggtacagatcagagcttagcccgctgctgctgtctacaacggagtggcaggta ttgccctgttccttcaccaccctaccggctctgtccactggtttgatccatctccatcagaacgtcgtggacgtacaatacctgtctgtgcctgcttgtggatgatgctgctgatagctcaagctgaggccgccctagagaacctggtggtcctcaacgcggcatcc gtggccggggcgcatggcattctctccttcctcgtgttcttctgtgctgctcggtacatcaagggcaggctggtccctggggc ggcatatgccctctacggcgtatggccgctactcctgctcctgctgcggttaccaccacgagcatacgccatggaccggga gatggcagcatcgtgcggaggcgcggttttcgtaggtctgatactcttgaccttgtcaccgcactataagctgttcctcgctag gctcatatggtggttacaatattttatcaccagggccgaggcacacttgcaagtgtggatcccccctcaacgttcgggggg geografic a test category at the content of the cateleaceceactgegggaetgggeceaeggggeetaegagaeettgeggtggeagttgageeegtegtettetetgatat gggggaggagatacatctgggaccggcagacagccttgaagggcagggtggcgactcctcgcgcctattacggccta ctcccaacagacgcgaggcctacttggctgcatcatcactagcctcacaggccgggacaggaaccaggtcgagggggag gtccaagtggtctccaccgcaacacatctttcctggcgacctgcgtcaatggcgtgtgttggactgtctatcatggtgccgg ctcaaagaccettgccggcccaaagggcccaatcacccaaatgtacaccaatgtggaccaggacctcgtcggctggcaag cgcccccggggcgcgttccttgacaccatgcacctgcggcagctcggacctttacttggtcacgaggcatgccgatgtcattccggtgcgccggcggcgacagcaggggggagcctactctcccccaggcccgtctcctacttgaagggctcttcgggcggtccactgctctgcccctcggggcacgctgtgggcatctttcgggctgccgtgtgcacccgaggggttgcgaaggcggtg gactttgtaccegtcgagtctatggaaaccactatgcggtccccggtcttcacggacaactcgtcccctccggccgtaccgc agacattccaggtggcccatctacacgcccctactggtagcggcaagagcactaaggtgccggctgcgtatgcagcccaa gggtataaggtgcttgtcctgaacccgtccgtcgccgccaccctaggtttcggggcgtatatgtctaaggcacatggtatcga ccctaacatcagaaccggggtaaggaccatcaccacgggtgcccccatcacgtactccacctatggcaagtttcttgccgac ggtggttgctctggggggcgcctatgacatcataatatgtgatgagtgccactcaactgactcgaccactatcctgggcatcgg cacagteetggaecaageggagaeggetggageggaetegtegtegeeaeegetaegeeteegggateggteaee gtgccacatccaaacatcgaggaggtggctctgtccagcactggagaaatccccttttatggcaaagccatccccatcgaga ccatcaagggggggaggcacctcattttctgccattccaagaagaaatgtgatgagctcgccgcgaagctgtccggcctcg

10

15

20

25

30

ggacccgaccttcaccattgagacgaccgtgccacaagacgcggtgtcacgctcgcagcggcgaggcaggactggt aggggcaggatgggcatttacaggtttgtgactccaggagaacggccctcgggcatgttcgattcctcggttctgtgcgagtgctatgacgcgggctgtgcttggtacgagctcacgccggcgagacctcagttaggttgcgggcttacctaaacacacag ggttgcccgtctgccaggaccatctggagttctgggagagcgtctttacaggcctcacccacatagacgcccatttcttgtcccagactaagcaggcaggagacaacttcccctacctggtagcataccaggctacggtgtgcgccagggctcaggctccacc tecategtgggaccaaatgtggaagtgteteataeggetaaageetaegetgeaegggecaaegeeetgetgtataggetg 10 gtcacgagcacctgggtgctggtaggcggagtcctagcagctctggccgcgtattgcctgacaacaggcagcgtggtcatt gtgggcaggatcatcttgtccggaaagccggccatcattcccgacagggaagtcctttaccgggagttcgatgagatggaa gagtgcgcctcacacctcccttacatcgaacagggaatgcagctcgccgaacaattcaaacagaaggcaatcgggttgctg caaacagccaccaagcaagcggaggctgctgctcccgtggtggaatccaagtggcggaccctcgaagccttctgggcga agcatatgtggaatttcatcagcgggatacaatatttagcaggcttgtccactctgcctggcaaccccgcgatagcatcactga 15 tggcattcacagcctctatcaccagcccgctcaccacccaacataccctcctgtttaacatcctgggggggatgggtggccgc ccaacttgctcctccagcgctgcttctgctttcgtaggcgccggcatcgctggagcggctgttggcagcataggccttggg aaggtgcttgtggatattttggcaggttatggagcaggggtggcaggcgctcgtgggcctttaaggtcatgagcggcgag atgccctccaccgaggacctggttaacctactccctgctatcctctcccctggcgccctagtcgtggggtcgtgtgcgcagc gatactgcgtcggcacgtgggccaggggggggggtgtgcagtggatgaaccggctgatagcgttcgcttcgcggggta 20 accaegate teccee acge act at gtg ect gag age get ge age acgtg teact caga tecte to tag tetta ecate actual content of the contcagetgetgaagaggetteaccagtggateaacgaggaetgeteeacgeeatgeteeggetegtggetaagagatgtitgg gattggatatgcacggtgttgactgatttcaagacctggctccagtccaagctcctgccgcgattgccgggagtcccctlcttc teatgteaacgtgggtacaagggagtetggeggggggacggcateatgcaaaccacetgeceatgtggagcacagateae cggacatgtgaaaaacggttccatgaggatcgtggggcctaggacctgtagtaacacgtggcatggaacattccccattaac 25 egtggaggttaegegggtgggggattteeactaegtgaegggeatgaeeactgaeaaegtaaagtgeeegtgteaggttee ggccccgaattetteacagaagtggatggggtgcggttgcacaggtacgctccagcgtgcaaacccctcctacgggagg aggtcacattcctggtcgggctcaatcaatacctggttgggtcacagctcccatgcgagcccgaaccggacgtagcagtgct 30 ggccagctcatcagctagccagctgtctgcgccttccttgaaggcaacatgcactacccgtcatgactccccggacgctgac ctcatcgaggccaacctcctgtggcggcaggagatgggcgggaacatcacccgcgtggagtcagaaaataaggtagtaat tttggactetttegageegeteeaageggaggaggatgagagggaagtateegtteeggeggagateetgeggaggteea ggaaattccctcgagcgatgcccatatgggcacgcccggattacaaccctccactgttagagtcctggaaggacccggactacgtecetecagtggtacaegggtgtecattgcegeetgecaaggeeeeteegataceaeeteeaeggaggaagaggaeg 35 gttgtcctgtcagaatctaccgtgtcttctgccttggcggagctcgccacaaagaccttcggcagctccgaatcgtcggccgt

5

10

15

20

25

30

cgacagcggcacggcaacggcctctcctgaccagccctccgacgacggcgacgcggggatccgacgttgagtcgtactcc tccatgccccccttgagggggagccgggggatcccgatctcagcgacgggtcttggtctaccgtaagcgaggaggctagtgaggacgtcgtctgctcgatgtcctacacatggacaggcgccctgatcacgccatgcgctgcggaggaaaccaagctgcccatca at gcactgag caact cttt gctccgt cacca caactt ggt ctat gctaca acatct cgcag cgcaag cct gcgg caact ctt gctacaacat ctt gagaagaaggtcacctttgacagactgcaggtcctggacgaccactaccgggacgtgctcaaggagatgaaggcgaaggc gtccacagttaaggctaaacttctatccgtggaggaagcctgtaagctgacgccccacattcggccagatctaaatttggct atggggcaaaggacgtccggaaacctatccagcaaggccgttaaccacatccgctccgtgtggaaggacttgctggaagacactgagacacca attgacaccaccat catggcaaaaaatgaggttttctgcgtccaaccagagaaggggggccgcaagccagetegeettategtatteecagatttgggggttegtgtgtgegagaaaatggeeetttaegatgtggteteeaeceteecteag gccgtgatgggctcttcatacggattccaatactctcctggacagcgggtcgagttcctggtgaatgcctggaaagcgaaga aatgccctatgggcttcgcatatgacacccgctgttttgactcaacggtcactgagaatgacatccgtgttgaggagtcaatct accaatgttgtgacttggcccccgaagccagacaggccataaggtcgctcacagagcggctttacatcgggggccccctga ctaattctaaagggcagaactgcggctatcgccggtgccgcggggggtgtactgacgaccagctgcggtaataccctca catgttacttgaaggccgctgcggcctgtcgagctgcgaagctccaggactgcacgatgctcgtatgcggagacgaccttgt cgttatctgtgaaagcgcggggacccaagaggacgaggcgagcctacgggccttcacggaggctatgactagatactctg cccccctggggacccgccaaaccagaatacgacttggagttgataacatcatgctcctccaatgtgtcagtcgcgcacg atgcatctggcaaaagggtgtactatctcacccgtgaccccaccaccacccctttgcgcgggctgcgtgggagacagctagac acactccagtcaattcctggctaggcaacatcatcatgtatgcgcccaccttgtggggcaaggatgatcctgatgactcatttctt ctccatccttctagctcaggaacaacttgaaaaagccctagattgtcagatctacggggcctgttactccattgagccacttgalung and the control of tcctacctcagatcattcaacgactccatggccttagcgcattttcactccatagttactctccaggtgagatcaatagggtggct tcatgcctcaggaaacttggggtaccgccttgcgagtctggagacatcgggccagaagtgtccgcgctaggctactgtcc caggggggaggaggctgccacttgtggcaagtacctcttcaactgggcagtaaggaccaagctcaaactccaatcccg gctgcgtccagttggattlatccagctggttcgttgctggttacagcgggggagacatatatcacagcctgtctcgtgcccga ttcctttctttctttctttggtggctccatcttagccctagtcacggctagctgtgaaaggtccgtgagccgcttgactgcagagagt gctgatactggcctctctgcagatcaagt

Other embodiments are within the following claims. While several embodiments have been shown and described, various modifications may be made without departing from the spirit and scope of the present invention.

WHAT IS CLAIMED IS:

1. A nucleic acid molecule comprising a region selected from the group consisting of:

a) an altered HCV NS3 encoding region coding for one or more NS3 mutations, wherein at least one of said NS3 mutations, identified by reference to the amino acid sequence numbering of SEQ. ID. NO. 1, is selected from the group consisting of:

amino acid 1095 being Ala,

amino acid 1202 being Gly, and amino acid 1347 being Thr;

b) an altered HCV NS5A encoding region coding for one or more NS5A mutations, wherein at least one of said NS5A mutations, identified by reference to the amino acid sequence numbering of SEQ. ID. NO. 1, is selected from the group

consisting of: amino acid 2041 being Thr,

a Lys insertion between residue 2039 and 2040.

amino acid 2173 being Phe,

amino acid 2197 being Phe,

amino acid 2198 being Ser,

30

amino acid 2199 being Thr, and

amino acid 2204 being Arg; and

- c) an altered encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) region containing one or more EMCV IRES mutations, wherein at least one of said EMCV IRES mutations, identified by reference to the nucleotide number of SEQ. ID. NO. 3, is an insertion at nucleotide 1736 of adenine.
 - 2. The nucleic acid molecule of claim 1, wherein said nucleic acid molecule comprises said NS5A encoding region.

3. The nucleic acid molecule of claim 2, wherein at least two of said NS5A adaptive mutations are present.

4. The nucleic acid molecule of claim 2, further comprising a region encoding for a HCV NS3 region, wherein said NS3 region may be the same or different than said altered NS3 region.

- 5. The nucleic acid molecule of claim 4, wherein said nucleic acid molecule is an HCV replicon comprising a HCV 5' UTR-PC region, said NS3 encoding region, an HCV NS4A encoding region, an HCV NS4B encoding region, said NS5A encoding region, an HCV NS5B encoding region, and a HCV 3' UTR.
- 10 6. The nucleic acid molecule of claim 5, wherein said HCV replicon further comprises a sequence encoding for a reporter protein.
 - 7. The nucleic acid molecule of claim 5, wherein said HCV replicon further comprises a sequence encoding for a selection protein.

15

8. The nucleic acid molecule of claim 5, wherein said HCV replicon further comprises a HCV core encoding region, a HCV E1 encoding region, a HCV E2 encoding region, a HCV p7 encoding region, and a HCV NS2 encoding region.

- 9. A nucleic acid molecule comprising a region selected from the group consisting of:
- a) an altered HCV NS3 encoding region containing one or more NS3 mutations, wherein at least one of said NS3 mutations, identified by reference to the nucleotide numbering of SEQ. ID. NO. 2, is selected from the group consisting of: nucleotide 3625 being cytosine, nucleotide 3946 being guanine, nucleotide 4380 being adenine,
- b) an altered HCV NS5A encoding region containing one or more 30 NS5A mutations, wherein at least one of said NS5A mutations, identified by reference to the nucleotide numbering of SEQ. ID. NO. 2, is selected from the group consisting of:
 - an insertion of 3 adenine residues between nucleotide 6458 and 6459, nucleotide 6463 being cytosine,
- 35 nucleotide 6859 being thymine or uracil,

nucleotide 6931 being thymine or uracil, nucleotide 6934 being cytosine, nucleotide 6936 being adenine, and nucleotide 6953 being adenine or guanine; and

- c) an altered encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) region containing one or more EMCV IRES mutations, wherein at least one of said EMCV IRES mutations, identified by reference to the nucleotide number of SEQ. ID. NO. 3, is an insertion at nucleotide 1736 of adenine.
- 10. The nucleic acid molecule of claim 9, wherein said molecule comprises said altered NS5A encoding region, and the nucleotide sequence of said altered NS5A region is provided for by bases 6258-7598 of SEQ. ID. NO. 2, or the RNA version thereof, modified with one or more of said NS5A modifications selected from the group consisting of:
- an insertion of 3 adenine residues between nucleotide 6458 and 6459, nucleotide 6463 being cytosine,
 nucleotide 6859 being thymine or uracil,
 nucleotide 6931 being thymine or uracil,
 nucleotide 6934 being cytosine,
 nucleotide 6936 being adenine, and
- 20 nucleotide 6936 being adenine, and nucleotide 6953 being adenine or guanine.
- The nucleic acid molecule of claim 10, wherein said molecule is an HCV replicon comprising a HCV 5' UTR-PC region, a modified HCV NS3-NS5B region, and a HCV 3' UTR, wherein said modified NS3-NS5B region comprises said altered NS5A region.
- 12. The nucleic acid molecule of claim 11, wherein said 5' UTR-PC region is the RNA version of bases 1-377 of SEQ. ID. NO. 2 and said 3' UTR is the RNA version of bases 9374-9605 of SEQ. ID. NO. 2.
 - 13. The nucleic acid molecule of claim 10, wherein said molecule is an HCV replicon comprising a HCV 5' UTR-PC region, a modified HCV NS3-NS5B region, and a HCV 3' UTR, wherein
- 35 said 5' UTR-PC region is the RNA version of bases 1-377 of SEQ. ID. NO. 2;

said 3' UTR is the RNA version of bases 9374-9605 of SEQ. ID. NO. 2; and said modified NS3-NS5B region consists of the RNA version of bases 3420-9371 of SEQ. ID. NO. 2 modified with one or more modifications selected from the group consisting of:

- nucleotide 4380 being adenine, nucleotide 3625 being cytosine, nucleotide 3946 being guanine, an insertion of 3 adenine residues between nucleotide 6458 and nucleotide 6459, nucleotide 6463 being cytosine,
- nucleotide 6859 being uracil,
 nucleotide 6931 being uracil,
 nucleotide 6934 being cytosine,
 nucleotide 6936 being adenine, and
 nucleotide 6953 being adenine or guanine.

- 14. The nucleic acid molecule of claim 13, wherein said replicon is a genomic replicon that further comprises the RNA version of nucleotides 378-3419 of SEQ. ID. NO. 2.
- 20 15. A nucleic acid molecule comprising the nucleic acid base sequence of bases 1-7989 of SEQ. ID. NO. 3, or the RNA version thereof, consisting of one or more different modifications selected from the group consisting of:
 - a) nucleotides 5335-5337 modified to code for arginine;
 - b) nucleotides 5242-5244 modified to code for phenylalanine;
- c) nucleotides 5314-5316 modified to code for phenylalanine;
 - d) nucleotides 5317-5319 modified to code for serine;
 - e) nucleotides coding for lysine inserted after nucleotide 4843;
 - f) nucleotides 2329-2331 modified to code for glycine, nucleotides 2764-2766 modified to code for threonine, nucleotides 5242-5244 modified to code for phanulal prime and an autra adaptating inserted after nucleotide 1736.
- 30 phenylalanine, and an extra adenosine inserted after nucleotide 1736;
 - g) nucleotides 4846-4848 modified to code for threonine, and nucleotides 5242-5244 modified to modified to code for phenylalanine;
 - h) nucleotides 4846-4848 modified to code for threonine, and nucleotides 5314-5316 modified to code for phenylalanine;

i) nucleotides 4846-4848 modified to code for threonine, and nucleotides 5317-5319 modified to code for serine;

- j) nucleotides 2329-2331 modified to code for glycine, and nucleotides coding for lysine inserted after nucleotides 4843;
- 5 k) nucleotides 5314-5316 modified to code for phenylalanine and nucleotides 5320-5322 modified to code for threonine;
 - l) nucleotides 4846-4848 modified to code for threonine, nucleotides 5314-5316 modified to code for phenylalanine, and nucleotides 5320-5322 modified to code for threonine;
- m) nucleotides 4846-4848 modified to code for threonine, nucleotides 5314-5316 modified to code for phenylalanine, and an extra adenosine inserted after nucleotide 1736; and
 - n) nucleotides 5314-5316 modified to code for phenylalanine, nucleotides 5320-5322 modified to code for threonine, and an extra adenosine inserted after nucleotide 1736;
- 15 and
 - o) nucleotides 5320-5322 modified to code for threonine.
 - 16. The nucleic acid of claim 15, wherein said one or more different modifications is selected from the group consisting of:
- 20 a) C5337A;
 - b) C5243T or U;
 - c) C5315T or U;
 - d) T or U5318C;
 - e) AAA inserted after 4843;
- 25 f) A2330G, G2764A, C5243T or U, and adenosine inserted 1736;
 - g) A4847C and C5243T or U;
 - h) A4847C and C5315T or U;
 - i) A4847C and T or U5318C;
 - j) A2330G and AAA inserted after 4843;
- 30 k) C5315T or U and G5320A;
 - 1) A4847C, C5315T or U, and G5320A;
 - m) A4847C, C5315T or U, and adenosine inserted 1736;
 - n) C5315T or U, G5320A and adenosine inserted 1736; and
 - o) G5320A.

17. The nucleic acid of claim 16, wherein said nucleic acid is RNA and comprises said nucleic acid base sequence.

- 5 18. The nucleic acid of claim 17, wherein said nucleic acid is RNA and consists of said nucleic acid base sequence.
- 19. An expression vector comprising a nucleotide sequence coding for the nucleic acid molecule of any one of claims 1-18, wherein said nucleotide
 sequence is transcriptionally coupled to an exogenous promoter.
 - 20. A recombinant cell human hepatoma cell, wherein said cell comprises the nucleic acid of any one of claims 5-8 and 11-18.
- 15 21. The recombinant cell of claim 20, wherein said hepatoma cell is an Huh-7 cell.
 - The recombinant cell of claim 20, wherein said cell is derived from a Huh-7 cell.
 - 23. A recombinant cell made by a process comprising the step of introducing into a human hepatoma cell the nucleic acid of any one of claims 5-8 and 11-18.
- 25 24. A method of making an HCV replicon enhanced cell comprising the steps of:

20

- a) introducing and maintaining a HCV replicon in a cell; and
- b) curing said cell of said HCV replicon to produce said replicon enhanced cell.
- 25. The method of claim 24, wherein said cell is a human hepatoma cell.
- 26. The method of claim 24, wherein said cell is a Huh-7 cell or is derived from a Huh-7 cell.

	27.	The method of claim 26, further comprising the step of
confirming the	ability	of said replicon enhanced cell to maintain an HCV replicon.

- 5 28 A method of making an HCV replicon enhanced cell containing a functional HCV replicon comprising the steps of:
 - a) introducing and maintaining a first HCV replicon in a cell;
 - b) curing said cell of said first replicon to produce a cured cell; and
- 10 c) introducing and maintaining a second HCV replicon into said cured cell, wherein said second HCV replicon may be the same or different than said first HCV replicon.
- The method of claim 28, wherein said cell is a human hepatoma cell.
 - 30. The method of claim 29, wherein said human hepatoma cell is a Huh-7 cell.
- 20 31. The method of claim 30, wherein said human hepatoma cell is derived from a Huh-7 cell.
 - 32. An HCV replicon enhanced cell made by the method of any one of claims 24-27.

33. An HCV replicon enhanced cell containing a HCV replicon made by the method of any one of claims 28-31.

- 34. A method of measuring the ability of a compound to affect HCV activity comprising the steps of:
 - a) providing said compound to the HCV replicon enhanced cell of claim 33; and
 - b) measuring the ability of said compound to effect one or more replicon activities as a measure of the effect on HCV activity.

35. The method of claim 34, wherein said compound is a ribozyme.

36. The method of claim 34, wherein said compound in an antisense nucleic acid.

- 37. The method of claim 34, wherein compound is an organic compound.
- 38. The method of claim 34, wherein said step (b) measures HCV protein production.
 - 39. The method of claim 33, wherein said step (b) measures production of RNA transcripts.

1	GCCAGCCCC GAT	TGGGGGC G	GACACTCCAC	CATAGATCAC	TCCCCTGTGA
51	GGAACTACTG TCT	TCACGCA G	GAAAGCGTCT	AGCCATGGCG	TTAGTATGAG
101	TGTCGTGCAG CCT	CCAGGAC C	CCCCCTCCC	GGGAGAGCCA	TAGTGGTCTG
151	CGGAACCGGT GAG	racaccg g	GAATTGCCAG	GACGACCGGG	TCCTTTCTTG
201	GATCAACCCG CTC	AATGCCT G	GGAGATTTGG	GCGTGCCCCC	GCGAGACTGC
251	TAGCCGAGTA GTG	TTGGGTC G	GCGAAAGGCC	TTGTGGTACT	GCCTGATAGG
301	GTGCTTGCGA GTG	CCCCGGG A	AGGTCTCGTA	GACCGTGCAC	CATGAGCACG
351	AATCCTAAAC CTC	AAAGAAA A	ACCAAAGGG	CGCGCCATGA	TTGAACAAGA
401	TGGATTGCAC GCAC	GTTCTC C	GGCCGCTTG	GGTGGAGAGG	CTATTCGGCT
451	ATGACTGGGC ACA	ACAGACA A	TCGGCTGCT	CTGATGCCGC	CGTGTTCCGG
501	CTGTCAGCGC AGGC	GCGCCC G	GTTCTTTTT	GTCAAGACCG	ACCTGTCCGG
551	TGCCCTGAAT GAAC	CTGCAGG A	CGAGGCAGC	GCGGCTATCG	TGGCTGGCCA
601	CGACGGGCGT TCCT	TGCGCA G	CTGTGCTCG	ACGTTGTCAC	TGAAGCGGGA
651	AGGGACTGGC TGCT	CATTGGG CO	GAAGTGCCG	GGGCAGGATC	TCCTGTCATC
701	TCACCTTGCT CCTG	CCGAGA A	AGTATCCAT	CATGGCTGAT	GCAATGCGGC
751	GGCTGCATAC GCTT	GATCCG G	CTACCTGCC	CATTCGACCA	CCAAGCGAAA
801	CATCGCATCG AGCG	AGCACG T	ACTCGGATG	GAAGCCGGTC	TTGTCGATCA
851	GGATGATCTG GACG	AAGAGC A	TCAGGGGCT	CGCGCCAGCC	GAACTGTTCG
901	CCAGGCTCAA GGCG	CGCATG C	CCGACGGCG	AGGATCTCGT	CGTGACCCAT
951	GGCGATGCCT GCTT	GCCGAA TA	ATCATGGTG	GAAAATGGCC	GCTTTTCTGG
1001	ATTCATCGAC TGTG	GCCGGC TO	GGGTGTGGC (GGACCGCTAT	CAGGACATAG
1051	CGTTGGCTAC CCGT	GATATT GO	CTGAAGAGC '	TTGGCGGCGA	ATGGGCTGAC
1101	CGCTTCCTCG TGCT	TTACGG TA	ATCGCCGCT (CCCGATTCGC	AGCGCATCGC
1151	CTTCTATCGC CTTC	TTGACG AG	GTTCTTCTG A	AGTTTAAACA	GACCACAACG
1201	GTTTCCCTCT AGCG	GGATCA AT	TTCCGCCCC '	тстссстссс	CCCCCCTAA
1251	CGTTACTGGC CGAA	GCCGCT TO	GGAATAAGG (CCGGTGTGCG	TTTGTCTATA ·
1301	TGTTATTTTC CACC	ATATTG CC	CGTCTTTTG (GCAATGTGAG	GGCCCGGAAA
1351	CCTGGCCCTG TCTT	CTTGAC GA	AGCATTCCT /	AGGGGTCTTT	CCCCTCTCGC
1401	CAAAGGAATG CAAG	GTCTGT TG	GAATGTCGT (GAAGGAAGCA	GTTCCTCTGG
1451	AAGCTTCTTG AAGA	CAAACA AC	CGTCTGTAG (CGACCCTTTG	CAGGCAGCGG
1501	AACCCCCAC CTGG	CGACAG GT	rgcctctgc (GGCCAAAAGC (CACGTGTATA

FIG. 1A

1551	AGATACACCT	GCAAAGGCGG	CACAACCCCA	GTGCCACGTT	GTGAGTTGGA
1601	TAGTTGTGGA	AAGAGTCAAA	TGGCTCTCCT	CAAGCGTATT	CAACAAGGGG
1651	CTGAAGGATG	CCCAGAAGGT	ACCCCATTGT	ATGGGATCTG	ATCTGGGGCC
1701	TCGGTGCACA	TGCTTTACAT	GTGTTTAGTC	GAGGTTAAAA	AACGTCTAGG
1751	CCCCCGAAC	CACGGGGACG	TGGTTTTCCT	TTGAAAAACA	CGATAATACC
1801	ATGGCGCCTA	TTACGGCCTA	CTCCCAACAG	ACGCGAGGCC	TACTTGGCTG
1851	CATCATCACT	AGCCTCACAG	GCCGGGACAG	GAACCAGGTC	GAGGGGGAGG
1901	TCCAAGTGGT	CTCCACCGCA	ACACAATCTT	TCCTGGCGAC	CTGCGTCAAT
1951	GGCGTGTGTT	GGACTGTCTA	TCATGGTGCC	GGCTCAAAGA	CCCTTGCCGG
2001	CCCAAAGGGC	CCAATCACCC	AAATGTACAC	CAATGTGGAC	CAGGACCTCG
2051	TCGGCTGGCA	AGCGCCCCC	GGGGCGCGTT	CCTTGACACC	ATGCACCTGC
2101	GGCAGCTCGG	ACCTTTACTT	GGTCACGAGG	CATGCCGATG	TCATTCCGGT
2151	GCGCCGGCGG	GGCGACAGCA	GGGGGAGCCT	ACTCTCCCCC	AGGCCCGTCT
2201	CCTACTTGAA	GGGCTCTTCG	GGCGGTCCAC	TGCTCTGCCC	CTCGGGGCAC
2251	GCTGTGGGCA	TCTTTCGGGC	TGCCGTGTGC	ACCCGAGGGG	TTGCGAAGGC
2301	GGTGGACTTT	GTACCCGTCG	AGTCTATGGA	AACCACTATG	CGGTCCCCGG
2351	TCTTCACGGA	CAACTCGTCC	CCTCCGGCCG	TACCGCAGAC	ATTCCAGGTG
2401	GCCCATCTAC	ACGCCCCTAC	TGGTAGCGGC	AAGAGCACTA	AGGTGCCGGC
2451	TGCGTATGCA	GCCCAAGGGT	ATAAGGTGCT	TGTCCTGAAC	CCGTCCGTCG
2501	CCGCCACCCT	AGGTTTCGGG	GCGTATATGT	CTAAGGCACA	TGGTATCGAC
2551	CCTAACATCA	GAACCGGGGT	AAGGACCATC	ACCACGGGTG	CCCCCATCAC
2601	GTACTCCACC	TATGGCAAGT	TTCTTGCCGA	CGGTGGTTGC	TCTGGGGGCG
2651	CCTATGACAT	CATAATATGT	GATGAGTGCC	ACTCAACTGA	CTCGACCACT
2701	ATCCTGGGCA	TCGGCACAGT	CCTGGACCAA	GCGGAGACGG	CTGGAGCGCG
2751	ACTCGTCGTG	CTCGCCACCG	CTACGCCTCC	GGGATCGGTC	ACCGTGCCAC
2801	ATCCAAACAT	CGAGGAGGTG	GCTCTGTCCA	GCACTGGAGA	AATCCCCTTT
2851	TATGGCAAAG	CCATCCCCAT	CGAGACCATC	AAGGGGGGA	GGCACCTCAT
2901	TTTCTGCCAT	TCCAAGAAGA	AATGTGATGA	GCTCGCCGCG	AAGCTGTCCG
2951	GCCTCGGACT	CAATGCTGTA	GCATATTACC	GGGCCTTGA	TGTATCCGTC
3001	ATACCAACTA	GCGGAGACGT	CATTGTCGTA	GCAACGGACG	CTCTAATGAC
3051	GGGCTTTACC	GGCGATTTCG	ACTCAGTGAT	CGACTGCAAT	ACATGTGTCA

FIG. 1B

3101	CCÇAGACAGT	CGACTTCAGC	CTGGACCCGA	CCTTCACCAT	TGAGACGACG
3151	ACCGTGCCAC	AAGACGCGGT	GTCACGCTCG	CAGCGGCGAG	GCAGGACTGG
3201	TAGGGGCAGG	ATGGGCATTT	ACAGGTTTGT	GACTCCAGGA	GAACGGCCCT
3251	CGGGCATGTT	CGATTCCTCG	GTTCTGTGCG	AGTGCTATGA	CGCGGGCTGT
3301	GCTTGGTACG	AGCTCACGCC	CGCCGAGACC	TCAGTTAGGT	TGCGGGCTTA
3351	ССТАААСАСА	CCAGGGTTGC	CCGTCTGCCA	GGACCATCTG	GAGTTCTGGG
3401	AGAGCGTCTT	TACAGGCCTC	ACCCACATAG	ACGCCCATTT	CTTGTCCCAG
3451	ACTAAGCAGG	CAGGAGACAA	CTTCCCCTAC	CTGGTAGCAT	ACCAGGCTAC
3501	GGTGTGCGCC	AGGGCTCAGG	CTCCACCTCC	ATCGTGGGAC	CAAATGTGGA
3551	AGTGTCTCAT	ACGGCTAAAG	CCTACGCTGC	ACGGGCCAAC	GCCCCTGCTG
3601	TATAGGCTGG	GAGCCGTTCA	AAACGAGGTT	ACTACCACAC	ACCCCATAAC
3651	CAAATACATC	ATGGCATGCA	TGTCGGCTGA	CCTGGAGGTC	GTCACGAGCA
3701	CCTGGGTGCT	GGTAGGCGGA	GTCCTAGCAG	CTCTGGCCGC	GTATTGCCTG
3751	ACAACAGGCA	GCGTGGTCAT	TGTGGGCAGG	ATCATCTTGT	CCGGAAAGCC
3801	GGCCATCATT	CCCGACAGGG	AAGTCCTTTA	CCGGGAGTTC	GATGAGATGG
3851	AAGAGTGCGC	CTCACACCTC	CCTTACATCG	AACAGGGAAT	GCAGCTCGCC
3901	GAACAATTCA	AACAGAAGGC	AATCGGGTTG	CTGCAAACAG	CCACCAAGCA
3951	AGCGGAGGCT	GCTGCTCCCG	TGGTGGAATC	CAAGTGGCGG	ACCCTCGAAG
4001	CCTTCTGGGC	GAAGCATATG	TGGAATTTCA	TCAGCGGGAT	ACAATATTTA
4051	GCAGGCTTGT	CCACTCTGCC	TGGCAACCCC	GCGATAGCAT	CACTGATGGC
4101	ATTCACAGCC	TCTATCACCA	GCCCGCTCAC	CACCCAACAT	ACCCTCCTGT
4151	ттаасатсст	GGGGGGATGG	GTGGCCGCCC	AACTTGCTCC	TCCCAGCGCT
4201	GCTTCTGCTT	TCGTAGGCGC	CGGCATCGCT	GGAGCGGCTG	TTGGCAGCAT
4251	AGGCCTTGGG	AAGGTGCTTG	TGGATATTTT	GGCAGGTTAT	GGAGCAGGGG
4301	TGGCAGGCGC	GCTCGTGGCC	TTTAAGGTCA	TGAGCGGCGA	GATGCCCTCC
4351	ACCGAGGACC	TGGTTAACCT	ACTCCCTGCT	ATCCTCTCCC	CTGGCGCCCT
4401	AGTCGTCGGG	GTCGTGTGCG	CAGCGATACT	GCGTCGGCAC	GTGGGCCCAG
4451	GGGAGGGGC	TGTGCAGTGG	ATGAACCGGC	TGATAGCGTT	CGCTTCGCGG
4501	GGTAACCACG	TCTCCCCCAC	GCACTATGTG	CCTGAGAGCG	ACGCTGCAGC
4551	ACGTGTCACT	CAGATCCTCT	CTAGTCTTAC	CATCACTCAG	CTGCTGAAGA
4601	GGCTTCACCA	GTGGATCAAC	GAGGACTGCT	CCACGCCATG	CTCCGGCTCG

FIG. 1C

4651 TGGCTAAGAG ATGTTTGGGA TTGGATATGC ACGGTGTTGA CTGATTTCAA 4701 GACCTGGCTC CAGTCCAAGC TCCTGCCGCG ATTGCCGGGA GTCCCCTTCT 4751 TCTCATGTCA ACGTGGGTAC AAGGGAGTCT GGCGGGGCGA CGGCATCATG 4801 CAAACCACCT GCCCATGTGG AGCACAGATC ACCGGACATG TGAAAAACGG 4851 TTCCATGAGG ATCGTGGGGC CTAGGACCTG TAGTAACACG TGGCATGGAA 4901 CATTCCCCAT TAACGCGTAC ACCACGGGCC CCTGCACGCC CTCCCCGGCG 4951 CCAAATTATT CTAGGGCGCT GTGGCGGGTG GCTGCTGAGG AGTACGTGGA 5001 GGTTACGCGG GTGGGGGATT TCCACTACGT GACGGCCATG ACCACTGACA 5051 ACGTAAAGTG CCCGTGTCAG GTTCCGGCCC CCGAATTCTT CACAGAAGTG 5101 GATGGGTGC GGTTGCACAG GTACGCTCCA GCGTGCAAAC CCCTCCTACG 5151 GGAGGAGGTC ACATTCCTGG TCGGGCTCAA TCAATACCTG GTTGGGTCAC 5201 AGCTCCCATG CGAGCCCGAA CCGGACGTAG CAGTGCTCAC TTCCATGCTC 5251 ACCGACCCT CCCACATTAC GGCGGAGACG GCTAAGCGTA, GGCTGGCCAG 5301 GGGATCTCCC CCCTCCTTGG CCAGCTCATC AGCTAGCCAG CTGTCTGCGC 5351 CTTCCTTGAA GGCAACATGC ACTACCCGTC ATGACTCCCC GGACGCTGAC 5401 CTCATCGAGG CCAACCTCCT GTGGCGGCAG GAGATGGGCG GGAACATCAC 5451 CCGCGTGGAG TCAGAAAATA AGGTAGTAAT TTTGGACTCT TTCGAGCCGC 5501 TCCAAGCGGA GGAGGATGAG AGGGAAGTAT CCGTTCCGGC GGAGATCCTG 5551 CGGAGGTCCA GGAAATTCCC TCGAGCGATG CCCATATGGG CACGCCCGGA TTACAACCCT CCACTGTTAG AGTCCTGGAA GGACCCGGAC TACGTCCCTC 5601 5651 CAGTGGTACA CGGGTGTCCA TTGCCGCCTG CCAAGGCCCC TCCGATACCA 5701 CCTCCACGGA GGAAGAGGAC GGTTGTCCTG TCAGAATCTA CCGTGTCTTC 5751 TGCCTTGGCG GAGCTCGCCA CAAAGACCTT CGGCAGCTCC GAATCGTCGG 5801 CCGTCGACAG CGGCACGGCA ACGGCCTCTC CTGACCAGCC CTCCGACGAC GGCGACGCGG GATCCGACGT TGAGTCGTAC TCCTCCATGC CCCCCCTTGA 5851 5901 GGGGGAGCCG GGGGATCCCG ATCTCAGCGA CGGGTCTTGG TCTACCGTAA GCGAGGAGGC TAGTGAGGAC GTCGTCTGCT GCTCGATGTC CTACACATGG 5951 ACAGGCGCCC TGATCACGCC ATGCGCTGCG GAGGAAACCA AGCTGCCCAT 6001 CAATGCACTG AGCAACTCTT TGCTCCGTCA CCACAACTTG GTCTATGCTA 6051 6101 CAACATCTCG CAGCGCAAGC CTGCGGCAGA AGAAGGTCAC CTTTGACAGA 6151 CTGCAGGTCC TGGACGACCA CTACCGGGAC GTGCTCAAGG AGATGAAGGC

FIG. 1D

6201	GAAGGCGTCC	ACAGTTAAGG	CTAAACTTCT	ATCCGTGGAG	GAAGCCTGTA
6251	AGCTGACGCC	CCCACATTCG	GCCAGATCTA	AATTTGGCTA	TGGGGCAAAG
6301	GACGTCCGGA	ACCTATCCAG	CAAGGCCGTT	AACCACATCC	GCTCCGTGTG
6351	GAAGGACTTG	CTGGAAGACA	CTGAGACACC	AATTGACACC	ACCATCATGG
6401	CAAAAAATGA	GGTTTTCTGC	GTCCAACCAG	AGAAGGGGG	CCGCAAGCCA
6451	GCTCGCCTTA	TCGTATTCCC	AGATTTGGGG	GTTCGTGTGT	GCGAGAAAAT
6501	GGCCCTTTAC	GATGTGGTCT	CCACCCTCCC	TCAGGCCGTG	ATGGGCTCTT
6551	CATACGGATT	CCAATACTCT	CCTGGACAGC	GGGTCGAGTT	CCTGGTGAAT
6601	GCCTGGAAAG	CGAAGAAATG	CCCTATGGGC	TTCGCATATG	ACACCCGCTG
6651	TTTTGACTCA	ACGGTCACTG	AGAATGACAT	CCGTGTTGAG	GAGTCAATCT
6701	ACCAATGTTG	TGACTTGGCC	CCCGAAGCCA	GACAGGCCAT	AAGGTCGCTC
6751	ACAGAGCGGC	TTTACATCGG	GGGCCCCCTG	ACTAATTCTA	AAGGGCAGAA
6801	CTGCGGCTAT	CGCCGGTGCC	GCGCGAGCGG	TGTACTGACG	ACCAGCTGCG
6851	GTAATACCCT	CACATGTTAC	TTGAAGGCCG	CTGCGGCCTG	TCGAGCTGCG
6901	AAGCTCCAGG	ACTGCACGAT	GCTCGTATGC	GGAGACGACC	TTGTCGTTAT
6951	CTGTGAAAGC	GCGGGGACCC	AAGAGGACGA	GGCGAGCCTA	CGGGCCTTCA
7001	CGGAGGCTAT	GACTAGATAC	TCTGCCCCCC	CTGGGGACCC	GCCCAAACCA
7051	GAATACGACT	TGGAGTTGAT	AACATCATGC	TCCTCCAATG	TGTCAGTCGC
7101	GCACGATGCA	TCTGGCAAAA	GGGTGTACTA	TCTCACCCGT	GACCCCACCA
7151	CCCCCTTGC	GCGGGCTGCG	TGGGAGACAG	CTAGACACAC	TCCAGTCAAT
7201	TCCTGGCTAG	GCAACATCAT	CATGTATGCG	CCCACCTTGT	GGGCAAGGAT
7251	GATCCTGATG	ACTCATTTCT	TCTCCATCCT	TCTAGCTCAG	GAACAACTTG
7301	AAAAAGCCCT	AGATTGTCAG	ATCTACGGGG	CCTGTTACTC	CATTGAGCCA
7351	CTTGACCTAC	CTCAGATCAT	TCAACGACTC	CATGGCCTTA	GCGCATTTTC
7401	ACTCCATAGT	TACTCTCCAG	GTGAGATCAA	TAGGGTGGCT	TCATGCCTCA
7451	GGAAACTTGG	GGTACCGCCC	TTGCGAGTCT	GGAGACATCG	GGCCAGAAGT
7501	GTCCGCGCTA	GGCTACTGTC	CCAGGGGGGG	AGGGCTGCCA	CTTGTGGCAA
7551	GTACCTCTTC	AACTGGGCAG	TAAGGACCAA	GCTCAAACTC	ACTCCAATCC
7601	CGGCTGCGTC	CCAGTTGGAT	TTATCCAGCT	GGTTCGTTGC	TGGTTACAGC
7651	GGGGGAGACA	TATATCACAG	CCTGTCTCGT	GCCCGACCCC	GCTGGTTCAT
7701	GTGGTGCCTA	СТССТАСТТТ	CTGTAGGGGT	AGGCATCTAT	СТАСТССССА

FIG. 1E

7751	ACCGATGAAC	GGGGAGCTAA	ACACTCCAGG	CCAATAGGCC	ATCCTGTTTT
7801	TTTCCCTTTT	TTTTTTTCTT	TTTTTTTTT	TTTTTTTTT	TTTTTTTTT
7851	TTCTCCTTTT	TTTTTCCTCT	TTTTTTCCTT	TTCTTTCCTT	TGGTGGCTCC
7901	ATCTTAGCCC	TAGTCACGGC	TAGCTGTGAA	AGGTCCGTGA	GCCGCTTGAC
7951	TGCAGAGAGT	GCTGATACTG	GCCTCTCTGC	AGATCAAGTA	CTTCTAGAGA
8001	ATTCTAGCTT	GGCGTAATCA	TGGTCATAGC	TGTTTCCTGT	GTGAAATTGT
8051	TATCAGCTCA	CAATTCCACA	CAACATACGA	GCCGGAAGCA	TAAAGTGTAA
8101	AGCCTGGGAT	GCCTAATGAG	TGAGCTAACT	CACATTAGTT	GCGTTGCGCT
8151	CACTGCCCGC	TTTCCAGTCG	GGAAACCTGT	CGTGCCAGCT	CCATTAGTGA
8201	ATCGTCCAAC	GCACGGGGAG	AGGCGGTTTG	CGTATTGGGC	GCACTTCCGC
8251	TTCCTCGCTC	ACTGACTCGC	TGCGCTCGTT	CGTTCGGCTG	CGGCGAGCCG
8301	TATCAGCTCA	CTCAAAGGCG	GTAATACGGT	TATCCACAGA	ATCAGGGGAT
8351	AACGCAGGAA	AGACCATGTG	AGCAAAAGGC	CAGCAAAAGG	CCAGGAACCG
8401	TAAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA	TAGGCTCCGC	CCCCTGACG
8451	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGGCGAAA	CCCGACAGGA
8501	CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC
8551	TGTTCCGACC	CTGCCGCTTA	CCGGATACCT	GTCCGCCTTT	CTCCCTTCGG
8601	GAAGCGTGGC	GCTTTCTCAT	AGCTCACGCT	GTAGGTATCT	CAGTTCGGTG
8651	TAGGTCGTTC	GCTCCAAGCT	GGGCTGTGTG	CACGAACCCC	CCGTTCAGCC
8701	CGACCGCTGC	GCCTTATCCG	GTAACTATCG	TCTTGAGTCC	AACCCGGTAA
8751	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG	GATTAGCAGA
8801	GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA
8851	CGGCTACACT	AGAAGGACAG	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG
8901	TTACCTTCGG	AAAAAGAGTT	GGTAGCTCTT	GATCCGGCAA	ACAAACCACC
8951	GCTGGTAGCG	GTGGTTTTTT	TGTTTGCAAG	CAGCAGATTA	CGCGCAGAAA
9001	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT	TTCTACGGGG	TCTGACGCTC
9051	AGTGGAACGA	AAACTCACGT	TAAGGGATTT	TGGTCATGAG	АТТАТСАААА
9101	AGGATCTTCA	CCTAGATCCT	AATTAAATTTA	AAATGAAGTT	ТТАААТСААТ
9151	CTAAAGTATA	TATGAGTAAA	CTTGGTCTGA	CAGTTACCAA	TGCTTAATCA
9201	GTGAGGCACC	TATCTCAGCG	ATCTGTCTAT	TTCGTTCATC	CATAGTTGCC
9251	TGACTCCCCG	TCGTGTAGAT	AACTACGATA	CGGGAGGGCT	TACCATCTGG

FIG. 1F

9301	CCCCAGTGCT	GCAATGATAC	CGCGAGAACC	ACGCTCACC	GCACCAGATT
9351	TATCAGCAAT	AAACCAGCCA	GCCGGAAGTG	CGCTGCGGA	AAGTGGTCCT
9401	GCAACTTTAT	CCGCCTCCAT	CCAGTCTATI	AGTTGTTGC	GGGAAGCTAG
9451	AGTAAGTAGT	TCGCCAGTCA	GCAGTTTGCG	TAACGTCGTT	GCCATAGCAA
9501	CAGGCATCGT	GGTGTCACGC	TCGTCGTTTG	GTATGGCTTC	ATTCAGCTCC
9551	GGCTCCCAAC	GATCAAGGCG	AGTTACATGA	TCCCCCATGT	TGTGCAAAAA
9601	AGCGGTTAGC	TCCTTCGGTC	CTCCGATCGT	TGTCAGAAGT	AAGTTGGCCG
9651	CAGTGTTATC	ACTCATGGTT	ATGGCAGCAC	TGCATAATTC	TCTTACTGTC
9701	ATGCCATCCG	TAAGATGCTT	TTCTGTGACT	GGTGAGTACT	CAACCAAGTC
9751	ATTCTGAGAA	TAGTGTATGC	GGCGACCGAG	TTGCTCTTGC	CCGGCGTCAA
9801	TACGGGATAA	TACCGCGCCA	CATAGCAGAA	CTTTAAAAGT	GCTCATCATT
9851	GGAAAACGTT	CTTCGGGGCG	AAAACTCTCA	AGGATCTTAC	CGCTGTTGAG
9901	ATCCAGTTCG	ATGTAACCCA	CTCGTGCACC	CAACTGATCT	TCAGCATCTT
9951	TTACTTTCAC	CAGCGTTTCT	GGGTGAGCAA	AAACAGGAAG	GCAAAATGCC
10001	GCAAAAAAGG	GAATAAGGGC	GACACGGAAA	TGTTGAATAC	TCATACTCTT
10051	CCTTTTTCAA	TATTATTGAA	GCATTTATCA	GGGTTATTGT	CTCATGAGCG
10101	GATACATATT	TGAATGTATT	TAGAAAAATA	AACAAATAGG	GGTTCCGCGC
10151	ACATTTCCCC	GAAAAGTGCC	ACCTGACGTC	TAAGAAACCA	TTATTACCAT
10201	GACATTAACC	ТАТАААААТА	GGCGTATCAC	GAAGCCCTTT	CGTCTAGCGC
10251	GTTTCGGTGA	TGACGGTGAA	AACCTCTGAC	ACTTGCAGCT	CCCGCAGACG
10301	GTCACAGCTT	GTCTGTAAGC	GGATGCCGGG	AGCAGGCAAG	CCCGTCAGGG
10351	CGCGTCAGTG	GGTGTTGGCG	GGTGTCGGGG	CTGGCTTAAC	TATGCGGCAT
10401	CAGAGCAGAT	TGTACTGAGA	GTACACCAGA	TGCGGTGTGA	AATACCGCAC
10451	AGATGCGTAA	GGAGAAAATA	CCGCATCAGC	CTCCATTCGC	CATTCAGACT
10501	CCGCAACTGT	TGGGAAGGGC	GGTCAGTACG	CGCTTCTTCG	CTATTACGCC
10551	AACTGGCGAA	AGGGGGATGT	GCTGCAAGGC	GATTAAGTTG	GGTAACGCCA
10601	GGGTTTTCCC	AATCACGACG	TTGTAAAACG	ACAGCCAATG	AATTGAAGCT
10651	TATTAATTCT	AGACTGAAGC	TTTTAATACG	АСТСАСТАТА	(SEQ. ID. NO.:3)

Fig. 1G

SEQUENCE LISTING

<110> Istituto Di Ricerche Di Biologia Molecolare P. Angeletti S.P.A. <120> HEPATITIS C VIRUS REPLICONS AND REPLICON ENHANCED CELLS <130> IT0003 PCT <150> 60/263,479 <151> 2001-01-23 <160> 13 <170> FastSEQ for Windows Version 4.0 <210>1<211> 3010 <212> PRT <213> Con 1 HCV isolate nucleic acid <400> 1 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn 1 5 10 Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly 20 25 30 Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala 35 40 45 Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro 50 60 Ile Pro Lys Ala Arg Gln Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly 65 70 75 80 Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp 85 90 95 Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 100 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys 115 120 125 Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu 130 135 140 Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp 150 155 Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile 165 170 175 Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Ile Pro Ala Ser Ala Tyr 180 185 190 Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp Cys Ser 195 200 205 Asn Ala Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro 210 215 220 Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val 225 230 235 240 Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr 245 250 255 Thr Ile Arg Arg His Val Asp Leu Leu Val Gly Ala Ala Ala Leu Cys

285

300

315

265 Ser Ala Met Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ala

280

295

310

Gln Leu Phe Thr Phe Ser Pro Arg Arg His Glu Thr Val Gln Asp Cys

Asn Cys Ser Ile Tyr Pro Gly His Val Thr Gly His Arg Met Ala Trp

260

Asp Met Met Met Asn Trp Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly Gly 370 375 380 Thr Tyr Val Thr Gly Gly Thr Met Ala Lys Asn Thr Leu Gly Ile Thr Ser Leu Phe Ser Pro Gly Ser Ser Gln Lys Ile Gln Leu Val Asn Thr Asn Gly Ser Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Asn Thr Gly Phe Leu Ala Ala Leu Phe Tyr Val His Lys Phe Asn Ser Ser Gly Cys Pro Glu Arg Met Ala Ser Cys Ser Pro Ile Asp Ala Phe Ala Gln Gly Trp Gly Pro Ile Thr Tyr Asn Glu Ser His Ser Ser 465 470 475 Asp Gln Arg Pro Tyr Cys Trp His Tyr Ala Pro Arg Pro Cys Gly Ile Val Pro Ala Ala Gln Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser 500 505 510 Pro Val Val Val Gly Thr Thr Asp Arg Phe Gly Val Pro Thr Tyr Ser Trp Gly Glu Asn Glu Thr Asp Val Leu Leu Leu Asn Asn Thr Arg Pro 530 540 Pro Gln Gly Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe 545 550 555 Thr Lys Thr Cys Gly Gly Pro Pro Cys Asn Ile Gly Gly Ile Gly Asn Lys Thr Leu Thr Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Thr Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Leu
595 600 605 Val His Tyr Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe 610 615 620 Thr Ile Phe Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Glu Ala Ala Cys Asn Trp Thr Arg Gly Glu Arg Cys Asn Leu Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu Leu Ser Thr Thr Glu Trp Gln Val Leu Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His Leu His Gln Asn Val Val Asp Val Gln Tyr Leu Tyr Gly Ile Gly Ser Ala Val Val Ser Phe Ala Ile Lys Trp Glu Tyr Val Leu Leu Leu Phe Leu Leu Ala Asp Ala Arg Val Cys Ala Cys Leu Trp Met Met Leu Leu Ile Ala Gln Ala Glu Ala Ala Leu Glu Asn Leu Val Val Leu Asn Ala Ala Ser Val Ala Gly Ala His Gly Ile Leu Ser Phe Leu Val Phe Phe Cys Ala Ala Trp Tyr Ile Lys Gly Arg Leu Val Pro
770 780 Gly Ala Ala Tyr Ala Leu Tyr Gly Val Trp Pro Leu Leu Leu Leu Leu Ala Leu Pro Pro Arg Ala Tyr Ala Met Asp Arg Glu Met Ala Ala

Ser Cys Gly Gly Ala Val Phe Val Gly Leu Ile Leu Leu Thr Leu Ser 825 820 Pro His Tyr Lys Leu Phe Leu Ala Arg Leu Ile Trp Trp Leu Gln Tyr 835 840 845 Phe Ile Thr Arg Ala Glu Ala His Leu Gln Val Trp Ile Pro Pro Leu 855 860 Asn Val Arg Gly Gly Arg Asp Ala Val Ile Leu Leu Thr Cys Ala Ile 870 875 His Pro Glu Leu Ile Phe Thr Ile Thr Lys Ile Leu Leu Ala Ile Leu 885 890 Gly Pro Leu Met Val Leu Gln Ala Gly Ile Thr Lys Val Pro Tyr Phe 905 900 910 Val Arg Ala His Gly Leu Ile Arg Ala Cys Met Leu Val Arg Lys Val 920 Ala Gly Gly His Tyr Val Gln Met Ala Leu Met Lys Leu Ala Ala Leu 930 935 940 Thr Gly Thr Tyr Val Tyr Asp His Leu Thr Pro Leu Arg Asp Trp Ala 950 955 His Ala Gly Leu Arg Asp Leu Ala Val Ala Val Glu Pro Val Val Phe 965 970 975 Ser Asp Met Glu Thr Lys Val Ile Thr Trp Gly Ala Asp Thr Ala Ala 980 985 990 Cys Gly Asp Ile Ile Leu Gly Leu Pro Val Ser Ala Arg Arg Gly Arg 995 1000 1005 Glu Ile His Leu Gly Pro Ala Asp Ser Leu Glu Gly Gln Gly Trp Arg 1015 1010 1020 Leu Leu Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu 1025 1030 1035 1040 1030 Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Arg Asn Gln Val Glu 1045 1050 1055 Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr 1060 1065 1070 Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys 1075 1080 1085 1075 1080 1085 Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val 1095 1090 1100 Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu 1105 1110 1115 112 1120 Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His 1125 1130 1135 Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu 1140 1145 1150Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro 1155 1160 1165 Leu Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val 1170 1175 1180 Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser 1190 1195 Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro 1205 1210 1215 Pro Ala Val Pro Gln Thr Phe Gln Val Ala His Leu His Ala Pro Thr 1220 1225 1230 Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly 1235 1240 1245 Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe 1250 1260 Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr 1265 1270 1275 1280 Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Ile Thr Tyr Ser Thr Tyr 1285 1290 1295 Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile 1305

Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly 1315 1320 1325 Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val 1330 1335 1340 Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro 1350 1355 1360 Asn Ile Glu Glu Val Ala Leu Ser Ser Thr Gly Glu Ile Pro Phe Tyr 1365 1370 1375 Gly Lys Ala Ile Pro Ile Glu Thr Ile Lys Gly Gly Arg His Leu Ile 1380 1385 1390 Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser 1395 1400 1405 Gly Leu Gly Leu Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser 1410 1415 1420 Val Ile Pro Thr Ser Gly Asp Val Ile Val Val Ala Thr Asp Ala Leu 1430 1435 1425 1440 Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr 1445 1450 1455 Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile 1460 1465 1470 Glu Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg 1475 1480 1485 Gly Arg Thr Gly Arg Gly Arg Met Gly Ile Tyr Arg Phe Val Thr Pro 1490 1500Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys **1510** 1515 1520 Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser
1525
1530
1535
Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln
1540
1540
1545
1550 Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile 1555 1560 1565 Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro 1570 1575 1580 Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro 1585 1590 1595 160 1600 Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro 1605 1610 1615 Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln 1620 1625 1630 Asn Glu Val Thr Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys 1640 1645 1635 Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly 1650 1655 1660 Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val 1665 1670 1675 1680 1680 Val Ile Val Gly Arg Ile Ile Leu Ser Gly Lys Pro Ala Ile Ile Pro 1685 1690 1695 Asp Arg Glu Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ala 1700 1705 1710 Ser His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe 1720 1715 1725 Lys Gln Lys Ala Ile Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu 1730 1735 1740 Ala Ala Ala Pro Val Val Glu Ser Lys Trp Arg Thr Leu Glu Ala Phe 1750 1755 1745 1760 Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala 1765 1770 1775 Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala 1780 1785 1790 1780 1785 1790 Phe Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln His Thr Leu Leu 1800

Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser 1810 1815 1820 Ala Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly 1830 1835 Ser Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly 1845 1850 1855 Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu 1860 1865 1870 Met Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser 1875 1880 1885 Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg 1895 1900 His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile
1905 1910 1915 192 1920 Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro 1925 1930 1935 Glu Ser Asp Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr 1940 1945 1950 Ile Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys 1955 1960 1965 Ser Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile 1970 1975 1980 Cys Thr Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu 1985 1990 1995 2000 Pro Arg Leu Pro Gly Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys 2005 2010 2015 Gly Val Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly 2020 2025 2030 Ala Gln Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly 2035 2040 2045 Pro Arg Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala 2050 2055 2060 Tyr Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg 2065 2070 2075 208 2080 Ala Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val 2085 2090 2095 Gly Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys 2100 2105 2110 Pro Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val 2115 2120 2125 Arg Leu His Arg Tyr Ala Pro Ala Cys Lys Pro Leu Leu Arg Glu Glu 2130 2140 2135 Val Thr Phe Leu Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu 2145 2150 2155 2160 Pro Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr 2165 2170 2175 Asp Pro Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg 2180 2185 2190 Gly Ser Pro Pro Ser Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala 2195 2200 2205 Pro Ser Leu Lys Ala Thr Cys Thr Thr Arg His Asp Ser Pro Asp Ala 2210 2215 2220 Asp Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn 2225 2230 2235 2240 2235 2240 Ile Thr Arg Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe 2245 2250 2255 Glu Pro Leu Gln Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala 2260 2265 2270 Glu Ile Leu Arg Arg Ser Arg Lys Phe Pro Arg Ala Met Pro Ile Trp 2275 2280 2285 Ala Arg Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro 2295

Asp Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ala Lys 2310 2315 2320 Ala Pro Pro Ile Pro Pro Pro Arg Arg Lys Arg Thr Val Val Leu Ser 2330 2325 2335 Glu Ser Thr Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe 2340 2345 2350 Gly Ser Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Ser 2355 2360 2365 Pro Asp Gln Pro Ser Asp Asp Gly Asp Ala Gly Ser Asp Val Glu Ser 2370 2375 2380 Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu 2385 2390 2395 2400

Ser Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val 2405 2410 2415 Val Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro 2420 2425 2430 Cys Ala Ala Glu Glu Thr Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser 2435 2440 2445 Leu Leu Arg His His Asn Leu Val Tyr Ala Thr Thr Ser Arg Ser Ala 2450 2455 2460

Ser Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp 2465 2470 2475 248 Asp His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr 2485 2490 2495 Val Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro 2500 2505 2510 Pro His Ser Ala Arg Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg 2515 2520 2525 Asn Leu Ser Ser Lys Ala Val Asn His Ile Arg Ser Val Trp Lys Asp 2530 2535 2540 Leu Leu Glu Asp Thr Glu Thr Pro Ile Asp Thr Thr Ile Met Ala Lys 2545 2550 2555 2560 Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala 2565 2570 2575 Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met 2580 2585 2590 Ala Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Ala Val Met Gly Ser 2595 2600 2605 Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val 2610 2615 2620 Asn Ala Trp Lys Ala Lys Lys Cys Pro Met Gly Phe Ala Tyr Asp Thr 2625 2630 2635 2649 Arg Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu 2640 2645 2650 2655 Ser Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile 2660 2665 2670 Arg Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser 2675 2680 2685 Lys Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu 2690 2695 2700 Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ala 2705 2710 2715 2720 Ala Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Cys Gly 2725 2730 2735 Asp Asp Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Glu 2740 2745 2750 Ala Ser Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro 2755 2760 2765 Pro Gly Asp Pro Pro Lys Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser 2770 2780

Cys Ser Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val 2790 2795 2800

```
Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp
                                       2810
                  2805
                                                            2815
 Glu Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile
             2820
                                 2825
                                                        2830
 Met Tyr Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe
         2835
                             2840
                                                 2845
 Phe Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys
     2850
                                               2860
                          2855
Gln Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln
2865
                     2870
                                          2875
                                                                2880
Ile Ile Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr
                 2885
                                     2890
                                                           2895
Ser Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly
             2900
                                2905
                                                     2910
Val Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala
         2915
                             2920
                                                   2925
Arg Leu Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu
                         2935
                                             2940
Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala
2945
                     2950
                                         2955
                                                                2960
Ala Ser Gln Leu Asp Leu Ser Ser Trp Phe Val Ala Gly Tyr Ser Gly
                2965
                                    2970
                                                           2975
Gly Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met
             2980
                                  2985
                                                       2990
Trp Cys Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro
         2995
                              3000
                                                   3005
Asn Arg
    3010
<210> 2
<211> 9605
<212> DNA
<213> Con 1 HCV isolate amino acid
<400> 2
gccagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactq
                                                                            60
tetteacgea gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae
                                                                           120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag
                                                                           180
gacgaccggg teetttettg gateaacceg etcaatgeet ggagatttgg gegtgeeece
                                                                           240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg
                                                                           300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac
                                                                           360
ctcaaagaaa aaccaaacgt aacaccaacc gccgcccaca ggacgtcaag ttcccgggcg gtggtcagat cgtcggtgga gtttacctgt tgccgcgcag gggccccagg ttgggtgtgc
                                                                           420
                                                                           480
gegegaetag gaagaettee gageggtege aacetegtgg aaggegaeaa eetateeeea
                                                                           540
aggetegeea geeegagggt agggeetggg eteageeegg gtaceeetgg eeeetetatg
                                                                           600
gcaatgaggg cttggggtgg gcaggatggc teetgteace eegtggetet eggeetagtt
                                                                           660
ggggccccac ggacccccgg cgtaggtcgc gcaatttggg taaggtcatc gataccctca
                                                                          720
cgtgcggctt cgccgatctc atggggtaca ttccgctcgt cggcgccccc ctagggggcg
                                                                          780
ctgccagggc cctggcgcat ggcgtccggg ttctggagga cggcgtgaac tatgcaacag ggaatctgcc cggttgctcc ttttctatct tccttttggc tttgctgtcc tgtttgacca
                                                                          840
                                                                          900
teccagette egettatgaa gtgegeaaeg tateeggagt gtaceatgte aegaaegaet
                                                                          960
gctccaacgc aagcattgtg tatgaggcag cggacatgat catgcatacc cccgggtgcg
                                                                         1020
tgccctgcgt tcgggagaac aactcctccc gctgctgggt agcgctcact cccacgctcg
                                                                         1080
cggccaggaa cgctagcgtc cccactacga cgatacgacg ccatgtcgat ttgctcgttg
                                                                         1140
gggcggctgc tctctgctcc gctatgtacg tgggagatct ctgcggatct gttttcctcg
                                                                         1200
tegeceaget gtteacette tegectegee ggeacgagae agtacaggae tgeaattget
                                                                         1260
caatatatcc cggccacgtg acaggtcacc gtatggcttg ggatatgatg atgaactggt cacctacagc agccctagtg gtatcgcagt tactccggat cccacaagct gtcgtggata
                                                                         1320
                                                                         1380
tggtggcggg ggcccattgg ggagtcctag cgggccttgc ctactattcc atggtgggqa
                                                                         1440
actgggctaa ggttctgatt gtgatgctac tctttgccgg cgttgacggg ggaacctatg
                                                                         1500
tgacaggggg gacgatggcc aaaaacaccc tcgggattac gtccctcttt tcacccgggt
                                                                         1560
catcccagaa aatccagctt gtaaacacca acggcagctg gcacatcaac aggactgccc
                                                                         1620
```

tgaactgcaa	tgactccctc	aacactgggt	tccttgctgc	gctgttctac	gtgcacaagt	1680
tcaactcatc	tagataccca	gaggggatgg	ccadetgead	CCCCatcgac	gcgttcgctc	1740
aggggtgggg	accepteect	tagaatgagt	. cacacacat		ccttattgtt	
49999t9999	geceateact	tacaatyayt	. cacacagett	gyaccayayc	Collarige	1800
ggcactacge	accceggeeg	tgcggtatcg	tacccgcggc	gcaggtgtgt	ggtccagtgt	1860
actgcttcac	cccaagccct	gtcgtggtgg	ggacgaccga	ccggttcggc	gtccctacgt	1920
acagttgggg	ggagaatgag	acqqacqtqc	tocttcttaa	caacacacaca	ccgccgcaag	1980
gcaactggtt	taactataca	togatgaata	acactagatt	Caccaagaga	tgcgggggcc	2040
cccatataa	cataggagg	ataaaaaata	geacegggee	caccaagacg	rgeggggee	
ccccgcgcaa	caccygygyg	accygcaaca	aaaccttgac	ergeeceaeg	gactgcttcc	2100
ggaagcaccc	cgaggccact	tacaccaagt	gtggttcggg	gccttggttg	acacccagat	2160
gcttggtcca	ctacccatac	aggctttggc	actacccctg	cactgtcaac	tttaccatct	2220
tcaaggttag	gatgtacgtg	gggggagtgg	agcacaggct	cgaagccgca	tgcaattgga	2280
ctcgaggaga	gcattataac	ctaggages	agazzagata	agaagttaga	ccgctgctgc	
tatataaaaa	gegetgeaac	ctggaggaca	gggacagacc	agayettaye	cegetgetge	2340
Lyccacac	ggagtggcag	gtattgeect	gttccttcac	caccctaccg	gctctgtcca	2400
ctggtttgat	ccatctccat	cagaacgtcg	tggacgtaca	atacctgtac	ggtatagggt	2460
cggcggttgt	ctcctttgca	atcaaatggg	agtatgtcct	attactcttc	cttcttctgg	2520
cggacgcgcg	catctatacc	tacttataga	tgatgctgct	gatageteaa	gctgaggccg	2580
ccctagagaa	cctaataatc	ctcaacacaa	catcoataac	caaaaaaaaa	ggcattctct	
cetteeteet	attattatta	cccaacgcgg	caccegege	cggggcgcac	ggcattetet	2640
	gricitate	getgeetggt	acatcaaggg	caggctggtc	cctggggcgg	2700
catatgccct	ctacggcgta	tggccgctac	tcctgctcct	gctggcgtta	ccaccacgag	2760
catacgccat	ggaccgggag	atggcagcat	cgtgcggagg	cgcggttttc	gtaggtctga	2820
tactcttgac	cttqtcaccq	cactataagc	tattecteac	taggeteata	tggtggttac	2880
aatatttat	caccadaacc	gaggagagagt	tacaaatata	dateccece	ctcaacgttc	2940
aaaaaaaaaa	castagagee	staataataa	catagagaga	gattettet	ctcaacytte	
9999999	cyatyccytt	accelectea	cgracycgar	ccacccagag	ctaatcttta	3000
ccaccaccaa	aatcttgctc	gccatactcg	gtccactcat	ggtgctccag	gctggtataa	3060
ccaaagtgcc	gtacttcgtg	cgcgcacacg	ggctcattcg	tgcatgcatg	ctggtgcgga	3120
aggttgctgg	gggtcattat	gtccaaatgg	ctctcatgaa	attaaccaca	ctgacaggta	3180
cqtacqttta	tgaccatctc	accccactgc	gggactgggc	CCSCGCGGGC	ctacgagacc	3240
ttacaataac	agttgagggg	atcatcttct	ctgatatgga	gaccaaggt	ataasstass	
aaacaaaaa	agergageee	geogeoetee	tegatatgga	yaccaayguu	accacciggg	3300
gggcagacac	cgcggcgtgt	ggggacatca	tcttgggcct	gcccgtctcc	gcccgcaggg	3360
ggagggagat	acatctggga	ccggcagaca	gccttgaagg	gcaggggtgg	cgactcctcg	3420
cgcctattac	ggcctactcc	caacagacgc	gaggcctact	tggctgcatc	atcactagcc	3480
tcacaggccg	ggacaggaac	caggtcgagg	gggaggtcca	agtggtctcc	accocaacac	3540
aatctttcct	auchacctuc	atcastage	tgtgttggac	tatatatat	actaccact	
Cannagaget	tassassass	gccaatggcg	tgtgttggat	Lycciaccac	ggrgccggct	3600
caaayacccc	tyccygecea	aagggcccaa	tcacccaaat	gtacaccaat	gtggaccagg	3660
acctcgtcgg	ctggcaagcg	cccccgggg	cgcgttcctt	gacaccatgc	acctgcggca	3720
gctcggacct	ttacttggtc	acgaggcatg	ccgatgtcat	tccggtgcgc	cggcggggcg	3780
acagcagggg	gagcctactc	tcccccaggc	ccgtctccta	cttgaagggc	tettegagea	3840
gtccactgct	ctacccctca	agacacacta	tagacatett	tcaaactacc	atatacecc	3900
gaggggttgg	assaccata	gggcacgccg	coatcacata	tatagaaaaa	gegegeacee	
gaggggttgc	gaaggeggeg	gactitgtac	ccgccgagtc	Latygaaacc	actatgeggt	3960
ccccggtctt	cacggacaac	tegteceete	cggccgtacc	gcagacattc	caggtggccc	4020
atctacacgc	ccctactggt	agcggcaaga	gcactaaggt	gccggctgcg	tatgcagccc	4080
aagggtataa	ggtgcttgtc	ctgaacccgt	ccgtcgccgc	caccctaggt	ttcggggcgt	4140
atatgtctaa	ggcacatggt	atcgacccta	acatcagaac	caaaataaaa	accatcacca	4200
cgggtgcccc	catcacutac	tccacctata	acaaatttet	taccascaat	gattactata	
addacaccts	taacataata	atatataata	gcaageeeee	caccaacage	ggttgtttg	4260
ggggcgccta	tyacattata	acacycyacy	agugudadud	aactgactcg	accactatcc	4320
tgggcatcgg	cacagteetg	gaccaagcgg	agacggctgg	agcgcgactc	gtcgtgctcg	4380
ccaccgctac	gcctccggga	teggteaceg	tgccacatcc	aaacatcgag	gaggtggctc	4440
tgtccagcac	tggagaaatc	cccttttatg	gcaaagccat	ccccatcgag	accatcaagg	4500
gggggaggca	cctcattttc	tgccattcca	agaagaaatg	tgatgagete	accacasaac	4560
tgtccggcct	cagactcaat	actataacat	attaccaaaa	cottactata	teagtastas	
Caactacca	agactcaat	geegeageae	actactgggg	ccccyacyca	teegecatae	4620
caactagcgg	agacgccacc	gccgcagcaa	cggacgctct	aatgacgggc	tttaccggcg	4680
atttcgactc	agtgatcgac	tgcaatacat	gtgtcaccca	gacagtcgac	ttcagcctgg	4740
acccgacctt	caccattgag	acgacgaccg	tgccacaaga	cacaatatca	cactcacaac	4800
ggcgaggcag	gactggtagg	ggcaggatgg	gcatttacag	atttataact	CCaddadaac	4860
ggccctcggg	catottooat	tectegatta	tatacasata	ctatoscoco	aactatactt	
dutaccases.	cacgeeegae	ananactor:	ttagggggg	cracyacycy	ggctgtgctt	4920
ggtacgagct	cacgeeegee	yayacctcag	ccaggetgeg	ggcttaccta	aacacaccag	4980
ggttgcccgt	ctgccaggac	catctggagt	tctgggagag	cgtctttaca	ggcctcaccc	5040
acatagacgc	ccatttcttg	tcccagacta	agcaggcagg	agacaacttc	ccctacctgg	5100
tagcatacca	ggctacqqta	tgcgccaggg	ctcaggetee	acctccatco	taggaccasa	5160
tgtggaagtg	totoatacgg	ctaaagccta	cactacecac	000000000	ctactatata	5220
auctagasac -	cattcasssc	gaggttacta	ccacacacac	gocaacycoc	tagataata	
ggctgggagc	agatasaat	gayyetatid	ccacacaccc	Cataaccaaa	cacaccatgg	5280
catgcatgtc	gyctyacctg	yaggicgica	cgagcacctg	ggtgctggta	ggcggagtcc	5340

tagcagctct	ggccgcgtat	tgcctgacaa	caggcagcgt	. ggtcattgtg	ggcaggatca	5400
tcttgtccgg	aaagccggcc	atcattcccg	acagggaagt	cctttaccgg	gagttcgatg	5460
agatggaaga	gtgcgcctca	cacctccctt	acatcgaaca	gggaatgcag	ctcgccgaac	5520
					gaggctgctg	5580
					catatgtgga	5640
					aaccccgcga	5700
					caacataccc	5760
					agcgctgctt	5820
					cttgggaagg	5880
					gtggccttta	5940
					cctgctatcc	6000
					cggcacgtgg	6060
	gggggctgtg					6120
						6180
	ccccacgcac					
	tcttaccatc					6240
	gccatgctcc					6300
	tttcaagacc					6360
	atgtcaacgt					6420
	atgtggagca					6480
tggggcctag	gacctgtagt	aacacgtggc	atggaacatt	cccattaac	gcgtacacca	6540
cgggcccctg	cacgccctcc	eeggegeeaa	attattctag	ggegetgtgg	cgggtggctg	6600
ctgaggagta	cgtggaggtt	acgcgggtgg	gggatttcca	ctacgtgacg	ggcatgacca	6660
	aaagtgcccg					6720
	gcacaggtac					6780
	gctcaatcaa					6840
	gctcacttcc					6900
	ggccagggga					6960
	cttgaaggca					7020
tcgaggccaa	cctcctgtgg	cggcaggaga	tgggcgggaa	catcacccgc	gtggagtcag	7080
	agtaattttg					7140
	tccggcggag					7200
	cccggattac			_		7260
	ggtacacggg					7320
	gaggacggtt					7380
	gaccttcggc					7440
	ccagccctcc					7500
	ccttgagggg					7560
	ggaggctagt					7620
	cacgccatgc					7680
	ccgtcaccac					7740
	ggtcaccttt					7800
	gaaggcgaag					7860
	gacgccccca					7920
	atccagcaag					7980
	gacaccaatt					8040
	ggggggccgc					8100
	gaaaatggcc					8160
	cggattccaa					8220
ggaaagcgaa	gaaatgccct	atgggcttcg	catatgacac	ccgctgtttt	gactcaacgg	8280
tcactgagaa	tgacatccgt	gttgaggagt	caatctacca	atgttgtgac	ttggcccccg	8340
	ggccataagg					8400
	gcagaactgc					8460
	taccctcaca					8520
	cacgatgctc					8580
	ggacgaggcg					8640
	ggacccgccc					8700
ccaatgtgtc	agtcgcgcac	gatgcatctg	gcaaaagggt	gtactatctc	acccgtgacc	8760
ccaccacccc	ccttgcgcgg	gctgcgtggg	agacagctag	acacactcca	gtcaattcct	8820
ggctaggcaa	catcatcatg	tatgcgccca	ccttgtgggc	aaggatgatc	ctgatgactc	8880
atttcttctc	catccttcta	gctcaggaac	aacttgaaaa	agccctagat	tgtcagatct	8940
acggggcctg	ttactccatt	gagccacttg	acctacctca	gatcattcaa	cgactccatq	9000
gccttagcgc	attttcactc	catagttact	ctccaggtga	gatcaatagg	gtggcttcat	9060

```
gcctcaggaa acttggggta ccgcccttgc gagtctggag acatcgggcc agaagtgtcc
                                                                         9120
gcgctaggct actgtcccag ggggggaggg ctgccacttg tggcaagtac ctcttcaact
                                                                         9180
 gggcagtaag gaccaagete aaacteacte caateeegge tgegteeeag ttggatttat
                                                                         9240
 ccagctggtt cgttgctggt tacagcgggg gagacatata tcacagcctg tctcgtgccc
                                                                         9300
 gaccccgctg gttcatgtgg tgcctactcc tactttctgt aggggtaggc atctatctac
                                                                         9360
 tccccaaccg atgaacgggg agctaaacac tccaggccaa taggccatcc tgtttttttc
                                                                         9420
9480
tectetttt tteetttet tteetttggt ggetecatet tagecetagt caeggetage
                                                                         9540
tgtgaaaggt ccgtgagccg cttgactgca gagagtgctg atactggcct ctctqcagat
                                                                         9600
caagt
                                                                         9605
<210> 3
<211> 10690
<212> DNA
<213> pHCVNeo.17 coding
<400> 3
gccagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg
                                                                           60
tetteaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae
                                                                          120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag
                                                                          180
gacgaccggg teetttettg gateaacccg etcaatgeet ggagatttgg gegtgeecce
                                                                          240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg
                                                                          300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac
                                                                          360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc
                                                                          420
eggeegettg ggtggagagg ctattegget atgaetggge acaacagaca ateggetget
                                                                          480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg
                                                                          540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc
                                                                          600
                                                                          660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga
                                                                          720
780
                                                                          840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg
                                                                          900
ccaggetcaa ggegegeatg eccgaeggeg aggatetegt egtgaeceat ggegatgeet
                                                                          960
gcttgccgaa tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc
                                                                         1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc
                                                                         1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc
                                                                         1140
agcgcatcgc cttctatcgc cttcttgacg agttcttctg agtttaaaca gaccacaacg gtttccctct agcgggatca attccgccc tctccctccc cccccctaa cgttactggc
                                                                         1200
                                                                         1260
cgaagccgct tggaataagg ccggtgtgcg tttgtctata tgttattttc caccatattg
                                                                         1320
ccgtcttttg gcaatgtgag ggcccggaaa cctggccctg tcttcttgac gagcattcct aggggtcttt cccctctcgc caaaggaatg caaggtctgt tgaatgtcgt gaaggaagca
                                                                         1380
                                                                         1440
gttcctctgg aagcttcttg aagacaaaca acgtctgtag cgaccctttg caggcagcgg
                                                                         1500-
aacccccac ctggcgacag gtgcctctgc ggccaaaagc cacgtgtata agatacacct
                                                                         1560
gcaaaggcgg cacaacccca gtgccacgtt gtgagttgga tagttgtgga aagagtcaaa
tggctctcct caagcgtatt caacaagggg ctgaaggatg cccagaaggt accccattgt
                                                                        1620
                                                                         1680
atgggatetg atetggggee teggtgeaca tgetttaeat gtgtttagte gaggttaaaa
                                                                        1740
aacgtctagg ccccccgaac cacggggacg tggttttcct ttgaaaaaca cgataatacc
                                                                         1800
atggcgccta ttacggccta ctcccaacag acgcgaggcc tacttggctg catcatcact
                                                                        1860
agcctcacag gccgggacag gaaccaggtc gagggggagg tccaagtggt ctccaccgca
                                                                         1920
acacaatett teetggegac etgegteaat ggegtgtgtt ggaetgteta teatggtgee
                                                                        1980
ggctcaaaga cccttgccgg cccaaagggc ccaatcaccc aaatgtacac caatgtggac
                                                                        2040
caggaceteg teggetggca agegeeece ggqqcqcqtt cettqacace atqcacetqc
                                                                        2100
ggcagctcgg acctttactt ggtcacgagg catgccgatg tcattccggt gcgccggcgg
                                                                        2160
ggcgacagca gggggagcct actetecece aggecegtet cetacttgaa gggetetteq
                                                                        2220
ggcggtccac tgctctgccc ctcggggcac gctgtgggca tctttcgggc tgccgtgtgc
                                                                        2280
accogagggg ttgcgaaggc ggtggacttt gtacccgtcg agtctatgga aaccactatg
                                                                        2340
cggtccccgg tcttcacgga caactcgtcc cctccggccg taccgcagac attccaggtg
                                                                        2400
geceatetae aegeceetae tggtagegge aagageaeta aggtgeegge tgegtatgea
                                                                        2460
gcccaagggt ataaggtgct tgtcctgaac ccgtccgtcg ccgccaccct aggittcggg
                                                                        2520
gcgtatatgt ctaaggcaca tggtatcgac cctaacatca gaaccggggt aaggaccatc
                                                                        2580
```

accacagggtg ecctapacat cataatatg gatgatggc actaactga cagtggttg 2700 atcctggggac cetapacat cataatatg gatgatggc actaactag accacat 2700 atcctggagac teggacagt cetagacaat cataatatg gatgatggc actagacaga actegtcgt 2700 atcctgcaca gatacgacat cetagacaga cetagacact 2820 gctctgtcac gatcagagagagagagagagagagagagagagagagagag							
tetggggggg ectatgacast cataatatgt gatgatgcc actoaactga etcgaccat 2700 atcstggcaceg etcagocatg cetggaccaa goggagacgg atcgetgtgtg gtctgtcac gacatggaga aatcccttt tatggcaaag catccaacat egaggaggtg gtctgtcac gacatggaga aatcccttt tatggcaaag catccaacat egaggaggtg gaggggggg geacetcat tttetgcat tecagagaag aatgtgatga getegecgg 2940 aagggggggg ggacactat tttetgcat tecagagaag aatgtgatga getegecgg 2940 aagctgtcac gecteggact caatgtgtga gcaatgacgt gggctttacc 3000 ataccaacta geggagacgt cattgtgta gcaacgacgac tectaatgac gggctttacc ggggatttc actcaactgat caacgacgaa gaagggcgt ggggacatcat actggggag atggggatta caaggacgt caatgtgcat caaggacgat caagtctagc ggggatttc actcacat tgagacgacg actggtcac aagacgggg gtcacqctag gaacggccc cgggaatgt cgattcctcg gttctgtgc acaggaggttgt accaacgac gacagggcgg gaggaacatct gattcctcg gttctgtgc agatgcttgt accaacgac gacaggacgac acquerccag actaaggacg agatgctta cacaacac accacataa accccattt cttgtcccaa acaaggacg agaggacaa ctcactcaca caaggttgc ccgtctgcca gacactctg gagttctgga gagagacaa ctcacctaa ctggtagact accaggctac gtgtgtgcc agggctcagg ctcacactc atcgtggac caatgtgga agtgttcat acggcaaag ctcagtgca accccactac atcgtggac atggagtga ggtgtgtca accaaggatg accacacaca accccataac caatggctga agtgtctaa acagggggg acaacacacacacacacacacacacac	accacgggtg	ccccatcac	gtactccacc	tatggcaagt	: ttcttgccga	cggtggttgc	2640
ctogcaccg ctacgcotcc gygateggte accgtgccac atccaaacat cgaggaggg ggctctgteca goactggaga aataccettt tattgcaat tccaagaaga aatgtgatga gctogccgg aagggggggg ggcacctcat tttetgcat tccaagaaga aatgtgatga gctogccgg 2940 aagsctgtcec goctoggact caatgctgta gcaatatacc ggggctttga tyataccgtc ggcgattte gatcagtgat cgatcgta gcaatatacc ggggcttgat tyataccgtc ggcgattte gatcagtgat cgaatgcaat acatgtgtca ccaagacagt cgacttcagc ggcggattte cagacagagagagagagagagagagagagagagagagag							
gctctytcca gcactggaga aatcccttt tatggcaaag catccccat cagagccatc 2880 aaggggggg ggacctcat tattctcat totatgcat troasqaaga aattyatgatga gctcyccgcg 2940 aagctytccg gcctcgagat caattgcty a gcatattace ggggccttga tytatccgtc 3060 gccgatttcg actcattgat ggactgat cattgtgta gcatattace ggggcctta caatcatgacgagagcattgat gattaccat gggagagag cattgtgata acatgtgtca cacaggaag tcacactag ggcgatttac catcagacagag actgacata acaggctty cacactagagag actcagatag acacgagag acacgagagagagagagagagagagag	atcctgggca	tcggcacagt	cctggaccaa	gcggagacgg	ctggagcgcg	actcgtcgtg	2760
aagggggggg ggcacctcat titetgcat tecasgaag aatgggatga gitageegg 2940 aagetgteeg geeteggat eaatgegta geatataee gggeettya tyateegte 3000 ataccaacta geggagaegt eattgtegta geaetgage citetaatgae gggettitee 3000 geggatitee acteggat egaetgeaat eaetgtgtea eecagaegt geaetteage 3120 etggaeegga cetteaceat tgagaegaag aeggtgeeat acagattigg daeteegga 3120 etggaeegga geaggaetgg taggaegaag aeggtgeeat acagattigg daeteeagga 3240 gaaeggeett eggeatgt gatteeteg gttetgtge agtgetatg eetaggetg 3240 gaaeggeette eggeatgt gatteeteg gttetgtge agtgetatg eetaggetg 3300 getiggtaeg ageegaett tutyteeag agtetgtggt teetaggetgt 3300 getiggtaeg ageegaett tutyteeag aggetagg etgagggetta cetaaacaca 3360 ceagggttge eegtegea ggaeeatetg gagtetggg aggagaeaa etteeetee 3420 accacatag aegeeattt tutyteeag aggetagg eeggagaeaa etceeteete 3420 accacatag aegeeattt tutyteeag attaaggetag etgagaeaa etceeteete 3420 caaatytgga atgteteea gggttgegee agggetagg etcaacaca etceetee atgtgggae caaatytgga agtgteteat aeggetaagg etcacacaca accecataac caaatacate atggeatgae tyteggetga eetggggt etcaacacaca accecataac caaatacate atggeatgae tyteggeeg gtattgeetg gtacagagae eetgggyet ggtaggegga atcatettgt eeggaaage ggcateatt eeggaeagg aagteetta eeggaggat gateeteetgag gaacaattea aacagaagge aateaggttg etgaaacagga aggtgetget ggtaggegga atcatettgt eeggaaage ggcateatt eeggaeagga gaggaggegg gagteetee gggaacatetg gagaetatta eagaeggaat eeggagtge teggaeacee teggeggatggggg accectegaag eettetggg aggetgegggatggggatggggggatgggggggggg	ctcgccaccg	ctacgcctcc	gggatcggtc	accgtgccac	: atccaaacat	cgaggaggtg	2820
aagctgtccg gectcggact cattgtgta gcatattacc ggggctttac 3000 ggcgatttcg actcagtgac cattgtgta gaacagaac ctctaatgac gggctttacc 3000 ggcgatttcg actcagtgtgt cattgtgta gaacagaacg actctaatgac gggctttacc 3000 ggcgatttcg actcagtgtgt cattcagtgac actggacgac acggacgacg cattgagacgacg acgggcatta cctaagacag 2310 caggatgag gaagacatgt tagaggcag atggcattt acaaggcttgt gattcagacg agaagacgag acggagacac actcaggagag ggaagacac caggagagaca ggtcagagacac caggagagaca actcagagagacacacggagacacacggacacacacacac							2880
ataccaacta geggaagegt cattgtcgta geaactggacg ctctaatgac gegetttacc 3120 ctggacccqa: ccttcaccat tgagacgacga accgtgccac aagacgcggt gractccaga 3120 ctggaccccqa: ccttcaccat tgagacgacga acggtgccac aagacgcgt gtcaccgcaga 3120 gaacggcag geagagacgg atggcattta acaggtttg gactccagag 3240 gaacggcacct cgggcatgtt cgattcctgg gtcttgtgtac acaggacgacga 3240 gaacggccct cgggcatgtt cgattcctag gtcttgtgtac acaggactacta acaggctac ggacatctag gagtcttaga gegggactat acaggacgac accaccatag acgccacttt ctttgccag acgatgctagg caggagacaac ctcagtaggt taccacacac 3360 ccaggagacat acaggctac ggtgtgcca agggctcagg ctccaccacca accactactaccacacacacacacaca	aaggggggga	ggcacctcat	tttctgccat	tccaagaaga	ı aatgtgatga	gctcgccgcg	2940
ggcgatttcg actcagtgat cgactgcaat acatgtgtca ccagacagt cgacttcagc ctgyacccopa ccttcaccat tyagacgacq acgtyccac aapacgcggt gtcacgctcg 3180 cagcggcgag gcaggactgt tagggcagt tagggcattt acaggtttg gcacgcccg gaacggccct cggcgatgtt cgattcctcg gttctgtgcg agtycatcgtc gcgcggcgt] 3300 gcttggtacg agtcacgcc cgccgagacc tcagttaggt tgcgggctta cctaaacaca 3360 ccagggttgc ccgtctycca ggaccatctg gagtcttgg agsocgctt tacacgccc accacata acgcccattt cttgtcccag actaagcagg cagggacaa cttccctac 3480 ctggtagacat accaggctac ggttytggc agggctacag cgccctgctg 3400 caatgtygga agtgtctcat acggctaaag cctacgtgc acgggcaac gccctgctg atatggctgg gagccgttca aaacgaggt actacacaca cacccataac caaatacatc atggctag gagccgttca aaacgaggt actacacaca cacccataac caaatacatc atggctag gagccgttca aaacgaggt actacacaca cacccataac caaatacatc atggcatgg agtcgtga cctggagtc gtcacgagca gcgtggtcat tgtgggcag gtcctagcag tcttggcgc gtattgcctg acacaagca gcgtggtcat tgtgggcag gtcatcatcttg ccggaaagc gccactaat cccgacagga agtccttta ccgggagtt gaacaattca aacaagagc accacaat cccacaaca acccataac agacgagggt gcgacaattca aacaagagc accaggtt ccaaacacc acccacaaca agacgagggt gctgctcccg tggtggaatc caagtggcg acctcacaac accacaaca agacgagggt gcggtttcccg tggtggaatc caagtggcg acctcacaacaca acccacacacacacacacacacaca							3000
ctygaccega cettcaccat tgagacgac accgtgccac aagacgcgt gtcacctcag 3240 gaacggccct cgggcatgt cgattcctcg gttctgtgcg agtgctatga gactccagag 3240 gaacggccct cgggcatgtt cgattcctcg gttctgtgcg agtgctatga cgcgggctgt 3300 gcttggtacg agetcacgcc cgccagacc tcagttaggt tgcgggctct cacaccca 3360 ccagggttgc ccgtctgcca ggaccatctg gagttctgg agagcgtctt tacaggccc 3420 accacatag acgccattt cttgtcccag actaagacgag caggagaca cttcccctca ctgggac accacaggagagagagagagagagagagagagag	ataccaacta	gcggagacgt	cattgtcgta	gcaacggacg	rctctaatgac	gggctttacc	3060
cagcgocca cagoactga tagggcatt acaggctttt acaggtttg gactccaga gaacggccct cagoactgt cattcctcg gittctgtagca agtycataga cagoggctg gcttgdtacg agtccaccc caccagacatt cagttaggt teggggctta cctaacaca a360 ccaggfttg ccgttcycca ggacactc gagttctgg agogcqtta tacacacaca a360 ccacgattg ccgttcycca ggacactcg agttctgga agogcqtta tacacacac accacatag accacatt cttgtccaca gactagaga caggaqacaa cttccctac accacatag accacatt cttgtccaca gactagaga caggaqacaa cttccctac ccggtagact accaggttca agggtgcaca gctaggtgc agggacaac ccagggtgt caaatgtgga aggtcttaa acggctaaag cctacgtgc acggacaac gcccttgtg acggacattacatcatg acggatgagat catacacac accacataac caaatacatc acgcatagca tytcggcg gattgcatgagagatgagagagtcattactctgggagagatcattcat							3120
gaacggccct cgggcatgtt cgattcctcg gttctgtog agtgctatga cgcgggctgt gcttggtacg agctcacqcc cgccqagacc tcagtaggt tgcgggctta cctaaacaca 3360 ccagggttgc ccgtctgcca ggaccattdt gagttctggg agagcgtctt cctaacacaca 3260 ccagggttgc ccgtctgcca ggaccattdt gagttctggg agagcgtctt tcacaggcta 3270 ccaactag agcccattd cttpcccaa actaagcagg cagagagcaac cttcccctca 3280 ctggtagcat accaggctac ggtgtgcgc caaatgtggg agtgcctta aaacggagat accacacaca acccataaac caaatacatc 3660 atggcatgca tttgggctgc gtattgcctg cctgagagc cctgggtgcd ggtaggcgga 3720 gtcctaagcag ctctggccgc gtattgcctg acaacacggag ggtggtcat ggtaggcgag 3720 gtcctaagcag ctctggccgc gtattgcctg acaacacggag aggtgtcttat tgtgggcagg 3780 gtcctaagcag ctctggccgc gtattgcctg acacacagga aggtgtctat tgtgggcagg 3780 gtcctaagcag ctctggcgcc acacacacacacacacacacacacacacacaca							3180
gettigstacg ageteacgee egeegagace teagtiagst begggetta ectaacaca 3360 accaegagite eegitetjeea ggacacatetig agatetetjeg agagegetet tacaegaea 3420 accaecatag acgeecatti ettgteecag actaegagg eagagaacaa etteecetaa 6420 accaecataga acgeecatti ettgteecag actaegagg eagagacaaa etteecetaa 3480 caaatagtgga agtgetetat acggetaaag ectaegetge acggeecaaa geeettgetggaa 3540 ataageetgga agateetteta acggetaaag ectaegetga ettgeggagga 3720 gteetagaetga tgteggetga actaegetga actaegetga ettggggetga ettgggaggt gtaecaegaea ettggggga 3720 gteetagaaga tgttggeega gtaetagaegggga atcaetatta ettgggeegg gtattgeeggga agteettatta eeggagagggg gtaecaeatt eeggaggggg agateettat eeggagggg 3780 ataatettg eeggaaagge ettgeaaacaag acgaggagt 3780 ataatettg eeggaaaggg ettgaaacaag ecaecaagaea agaggagtet 3960 getggeeteeg ggtggaate caaatattaa gaaggggga accetegaag accetegaag ecaecaeaca agaggagget 3960 getggeeteeg ggtggaate eagaggggg accetetate ecaecagagaa ecetetegag gaagaetatat 4020 getggatatta tacaegagga acaetatta gaaggettgt ecaetetgee tggeaacaec 4080 gegatagatat tacaeagae ettaacaeca geeegetaa eaceecaacat 4140 gegatagata eaceetagagga ettaacaeca geeegetaa eaceetagaggeeggggggggggggggggggggggggggg							
ccagaggitige cegicitycea ggaccattet diffeccaga acatagcagg cagagagaca ettecetae 3480 ciggitageat accagcatte diffeccaga acctagcagg cagagagacaa ettecetae 3480 ciggitageat accagegatae ggatyagege agggsteagg clecacetee ategitygga aggiteteat acggstaagg catacagetge agggsteagg gaccateggit actagityga aggiteteat acggatyagg cataggity actagityga aggiteteat acggatyga citygaggity agggsteggityga gaggatygaggityga actagitygaggitygaggitygaggaggaggateaggytetagaggagaggateaggytetagaggagagateaggatygatygagagagagatygatygagagagagagaga							
accecacatag acgeceattt cttgtcccag actaagcag caggagacaa cttcccctac ctgtgtagcat accaggacta gdtgtcgcca agggctcag ctcacactce atcgtggac 3540 caaattgtgga agtgtctcat acggctaaag cctacgctgc agggccaac gccctgctg 3600 tataggctgg gagcgttca aaacaggtt actaccacca cacccataac caaatcact 3660 atggcatga tgtcggctgg cttgtgggggg cttgggggggggg	gcttggtacg	agctcacgcc	cgccgagacc	tcagttaggt	tgcgggctta	cctaaacaca	
ctgqtaqcat accaqgetac ggtqtgcgcc agggctagg ctcactcc atcgtggga 3540 tataggctgg gagccgttca aaacggagta accacgctgc acgggcacac gccctqctg 3600 tataggctgg gagccgttca aaacgaggtt actacccaca accccataac caaatacatc 3660 atggatggat gtcgtggatgga ctcggaggatgcgtcatcagcagga ctctgggtgct tgtggggggg 3720 gtcctagcag ctctgggtgct tcgggggggggatgcgc gtcacacactc cccgacagga agtccttta ccggaggtt 3840 gacagatgag aggggggggggaggggggaggagggggagggggggg							
caaatgtgga agtgteteat acggetaaag ectacgetge acgggecaac gecettgetg 3600 atggegtgg ggcogttea aaacgaggtt actacgacac accectataac caaatacatc afgeatgeatgeat tgteggetge cttggaggte gtcacgagca cctgggtget ggtaggegga 3720 gtcetageag ctctggcege gtattgeetg acaacaaggca gegtggteat tgtggagggag 3780 atcatettgt ceggaaagce gccatcatt cccgacagga agtecttaa ccgggaggte 3840 gatgagatgg aagagtettea acaacgagagg cctacaaccet ccttacateg aacaagggaa gagtgetge 3900 getggteetg aagagagage cgacaaatte caacagagag aagagagag aggeggaggg 3960 getggteetg tggtggaate caagtgggtgg accetcagaag ccttcaggga gagagaggggggggggggggg							
tataggctgg aggccgttca aaacgaggtt actaccacac accccataac caaatacatc ataggcatgga tyggctggtgg gtaggtggg 3720 gtcctaggag ctctggccg gtattgctg acaacaggca actgggtgat tygggaggg 3720 gtcctagaga ctctggccgc gtattgctg acaacaggca gcgtggtcat tygggaagg 3780 atcatcttg ccggaaagc ggccatcatt cccgaaagga agtccttta ccgggagtg 3840 gatgagatgg aagagtggc accacaacaga acacaggaa agcgcgtcc 3900 gaacaattca acacaggag aatcgggttg ctgcaaacag caccaagaa agcggaggc 3900 gaacaattca tcagggagt acaatattta gcaggattg cccccaagaa agcggaggct 3960 gcgctcccc tygtggaac caatggggg acccccgaaga ccttctggc gaagcatatg 4020 tygaattca tcactgatgga atcacacaga cccccaacagaa agcggaggct taggaattcac tagggagat acaatattta gcaggcttg ccatctctgc tygcaacccc 4080 gcgtatctcgt ttaacatcct gggggggggggggggg	ciggtageat	accaggetae	ggtgtgcgcc	agggctcagg	ctccacctcc	atcgtgggac	
atggcatgca tgtcggctgg ctattgctg gcatgcgcggggaggaggcgggggggggg							
gtctctagcag ctcttggccg gtattgctg acaacaggca gcgtggtcat tggggagttc gatgagatgg aagagtggc ctcacactc ccttacatcg aacagggaat gcagctcgc 3900 gaacaattca aacagaaggc atcacactc ccttacatcg aacagggaat gcagctcgc 3900 gaacaattca aacagaaggc atcacactgag agtcacacaca agcggagggc 3960 gctgctcccg tggtggaatc caattatta gcagcttgt ccactctgcc tggcaacccc 4080 gcgatagcat cactgatgga atcacacacacacacacacacacacacacacacacaca							
atcatcttgt cegaaagec ggccatcatt cccgacagg aagtccttta ccggaagtc gatgagatgg aagagtggc ctcacacctc ctttacatg accaggaat gaagtcgcc 3900 gaacaattca aacagaagg aatcgggtt ctgcaaacag ccaccaagca agcggaggct 3960 gctgctcccg tggtggaat caagtagggag acccttcgaag ccatctage gaagcatatg 4020 tggaatttca tcagcgggat acaatattta gcaggcttgt ccactctgcc tggacacccc 4080 gcgatagcat cactgatggc attacacagc tctatcacca gcccgctcac cacccaacat 4140 accctctgt ttacactcc gggggatgg gtggcgcc aacttggtcc tccactggt 4260 gcttctggtt tcqtaagcgc cggcatcgct ggagcggcg ttggcagcat aggccttggc 4200 gcttctggtt tcqtaagcgc cggcatcgct ggagcagggg tggcaggcg gcttgggcc 4200 aaggtgtgt tggatattt ggacaggtat ggagcagggg tggcaggcg gcttgggcc 4320 tttaaggtca tggacggcg aatgccctcc accgaagac tggttacact actccctgct 4380 atcctctccc ctggcgccc agtgctcgg gatgcgcgg gtcgggggcg gctgggggc gtggaggcca gtggggggcg gtggagggcc gtgtgggcca atgcgccagaggagggggggggg	atggtatgta	ctctggctga	atattagata	gccacgagca	cctgggtgct	ggtaggcgga	
gatgaqattga aagaqtgcgc ctcacacctc ccttacatcg aacagggaat gcagctcgcc 3900 gaacaattca aacaggaaggc aatcgggttgt ctgcacacag ccccaagca agggaggct 3960 gctgctcccg tggtggaatc caagtggcgg accctcgaag ccttcttggc gaagcatatg 4020 tggaattca tcagcgggat acaatattta gcaggcttgt ccactctgcc tggcaacccc 4080 gcgataagcat cactgatggc attacacagc tctatcacca gccgctcac cacccaacat 4140 accctcctgt ttaacatcct gggggggtgg gtggccgcc aacttgctcc tcccaagcgt 4200 gcttctgctt ttagctgggcg gatgccgtg ggaggggtg ttggcaggcg cctcgtggc 4200 gcttctgtt ttaaggtcac gggaatcatt ggaggaggtg ttggcaggcg gtggcagggg gtgtctgtgtgg ggagggggg gtggcggga gatgcctcc accgaagacc tggttaacct actcctgt 4380 atcctctccc ttggcgcct agtcgtggg gtgtgtggg gtgtgtggg ggaggatcat ggctgggac 4440 gtggggcccag ggaaggggg tgtgcagtgg atgacacag tggttaacct actcctgct 4380 gtgaggccag ggagggggg gtgtgggg gtgtgtggg agggatcat ggctgtggac 4440 gtggggcccag ggaaggggg tgtgcagtgg atgacacaga tcgctccacac tggtcacaca 4560 cagatctct ctagtcttac catcactcag ctgctgaaga ggcttcacca gtggtacac 4560 cagatctct ctagtcttac catcactcag ctgctgaaga ggcttcacca gtggatcac 4620 gaggactgt cacaccatgtgtac cacaccatggtc cagtcaagag ggcttcacca gtggatcac 4620 gaggacttgt ccacaccatg ctccgcacat gtggtaaagag aggtttggga ttgggatatgc 4680 acggtattga ctgattcaca acgtgggtac aggcaagagtc tggcaagag atgttggga ttggatatgc acggaactgt tccacactggg 4740 gtcccttct tctcatgtca acgtggtac aaggaggtct ggcgaggca ttgggaggg 4740 gtcccttct tctcatgtca acgtggtac aaggaggtct ggcgagggg ttggggggg 4740 gtcccttct tctcatgtca acgtgggtac aaggaggata tccacacgggc ctggagacatg ggacacaaga cacggacatg gaaaaacgg 4800 accacagggc ctggagactgg ggttacgcag gtgtagggaga atccccatacgg ggaggggg 4800 accacagggc ctggagacatga ggttacggag gtgtaggggg ggttgagaa actccccatacgg gaggggg 4800 accactggag acgtagagagg ggttaggggagacacaacactgaagagagagagagagaga	atcatcttct	ccccggeege	grantgeerg	acaacaggca	gegtggteat	tgtgggcagg	
gacacattica acagaagg aategggttg etgeaacag caccaagaa ageggagget 3960 getgeteceg tggtgaate caagtggegg accetectage cettetingge gaageatatg 4020 tggaattica teagegggat acaatattia geaggetigt ecactetinge tggeaacee 4080 gegatageat cactgatgge atteacagee tetateacea geeegeteae caccaacat 4140 accetectity thacaateet ggggggatgg gtggeegee aactgetee tecagagget 4200 gettetiget tegtagegee eggaateget gagaeggege tiggeaggaa aggeetiggg 4200 aaggtgetit tiggaaggat tegtaggeggegee acttysteet tecagagget 4200 aaggtgetit tiggaaggaa gatgeetit ggaaggagge gettigggee 4260 aaggtgetit tiggaaggaa gatgeetite acgaaggat tiggaaggag tiggaaggagge gettigggee 4320 titaaggtea tiggaaggagg gatgetitigg ageaggagge gettigggee 4320 titaaggtea tiggaaggagg gatgetitigg aggaaggagg gettiggagge 4320 titaaggtea tiggaaggagg gatgetitigg aggaaggagg eggtiggaaca 4480 giggageeca gggaaggagge tigtgaaggaggaggaggaggee 200 giggaggaggagge 200 giggaggaggaggaggaggaggaggaggaggaggaggagg	gatgagatgg	aagaatacac	ctcacacctc	ccttacatca	aagttttta	ccgggagttc	
gctgatacca toagcaga acaatattta gcaggettg ccactcagac gaagcatatg 4020 gcgatagcat toagcgggat acaatattta gcaggettgt ccactcagc tggacacccc 4080 gcgatagcat cactgatgg attoacagc tctatcacca gcccgctcac caccaacat 4140 accctcctgt ttaacatct ggggggatgg gtggccgcc aacttgctc tccaagcgt 4200 gcttctgct ttagtgcg cggaatcgt ggagcaggg tgggcgcgc tggcaggggg aggccttggg 4260 aaggtgttg tggatatttt ggaagtata ggagcaggg tggcagggg gcctgggg 4260 aaggtgttg tggatatttt ggaagtata ggagcagggg gtggcggg gtggagggg gtggagggg ggaggggg gagggggg gagggggg gtggagggg gtggagggg gggagggg gggagggg gtggagggg gggagggg gggaggggg ggggggg gtggagggg gtggagggg gggagggg gggagggg 4380 gggaaccacg gggaaggggg tggaaccag gggaagggg tggaagggg gtggaagggg gggaggggg gggaagggg 4400 ggggaaccag gggaagggg gggggggg gggaagggg gggaagggg 4500 gggaaggaggg gggaaggggg ggggggggg gggggg							
ggatatttca cacegaggat acaatattta gcaggcttgt cacetctgcc tggcaacccc gcgtataggat tacacagcc tctatacaca gcccgctcac cacccaacat accctcctgt ttaacatct ggggggatgg gtggccgcc acttyccctct tcccaacggt gcttctgctt tcgtaggcc cggcatcgct ggagcggctg ttggcagcat aggccttggg 4260 aaggtgcttg tggatatttt ggcaggttat ggagcagcgg ttggcagcac actcctgct 4380 atcctctccc ctggcgccc agtcgctcgg gtegtgtgcc caccgagacc tggtgacccc gtgggacaggggg attcacacgg gtegtgtgcc caccgagacct actccctgct 4380 atcctctccc ctggcgccc agtcgtcggg gtegtgtgcc caccgagacct ggctgcagca gtgggggcaag ggaggggc tgtgcagtgc caccgagacct gtgtcaccc gtgggacacagg tctcccccac gcactatgg ctggagacg acgttcacc actcctctct ggggacacacg tctcccccac gcactatgg ctggagaggcg aggttcaccac gtggatacac 4500 ggtaaccacg tctcccccac gcactatgg ctggagagg aggttcaccac gtggatcacac gaggactct caccaccatg ctccggctcg tggctaagag acgttgcacc acgtgtacac 4500 gaggactgct caccaccatg ctccggctc tggctaagag ggtttcacca gtggatcacac gaggactgtc ccacccatg ctccggctcg tggctaagag acgttcacac acgtgaatcac 460 gatcccttt tctatgtca acctggctc cagtcaagag atgtttggga ttggatatgc 4680 accggaggggc ctggagaggaggt aggagggggggggg							
gcgatagcat cactgatggc attcacagcc tctatcacac gcccgctcac caccacaat 4140 accotcctgt tracatcct gggggatgg gtggccgcc aacttgctc tcccagggct 4200 gcttctgctt tcgtaggcc cggcatcgct ggagcggctg ttggcaggca aggccttggg 4260 aaggtgcttg tggaagttat ggcaggtgat ttggcaggcg gtcgttggc 4320 aaggtgcttg tggaaggtca tggacagggg ttggtaggcc gtcgtggcc 4320 attcactctcc ctggcgccct agtcgtcgg gtcgttgtgc cagcgatact gcgtcggcac 4440 gtggggccag gggaggggg tgtgtgggg tggtagggcg cactgtgcgc 4400 ggtaaccacg tcccccaca gcactattgt cctgagagca cagctgcaca acgctgcaca cagatcctct ctagtcttac catcactcag ctgctgaaga ggcttcacca gtggatcaca 4600 gagagactgct ccacecaca gcactattgt cctgagaga acgctgcaca acgtgtacac 4600 gagagactgct ccacecaca gcactattgt cctgagaga ggcttcacca gtggatcaca 4600 gagagactgct ccacecaca gacctggctc tggctaagaga atgtttggga ttggatatag 4600 caaaccacct tctattgta acgtggtac aaggagatct ggcgaggaga cggcatcatg 4800 caaaccacct gcccatgtgg acgagactgt tggcaagga acgagactgg 4800 accacgggcc ctaggacct tagtaacacag tggcatggaa cattccccat taacgcgtac 4920 accacgggcc ctggcaagc ctccccagg ggtacaggg gttcgggggat tccacagggcg ggtaggggc gcgcatcatg 4800 accactgaga gatacgtga ggtacacgg ggtaggggat tccacagggcg ggtaggggg gttccacagggcc ctgcacagggc ctgcaattat ctagggggc ggtaggggg ggtagggggat accacagggcc ctgcaacggc ctccccaggg ggtagggggat tccacacagggcc ggtaggggat 4920 accacatgga ggtacaggg ggtacaggg ggtagggggat tccacacagggcc ggcatcatag 4920 accactgaag agtacgtga ggtacaggg ggtaggggat tccacacaggagggt 5040 accactgaag ggttgcaag ggtacacag ggtggagaacacacacacacacacacacacacacacacac	togaatttca	tcagcgggatt	acaatattta	acadacttat	ccactctgcc	tagcacacg	
acctcctctt tegtaggee eggeateget ggageggetg ttggeagee acttgete tectaggege 4260 aagtgettt tegtaggee eggeateget ggageggetg ttggeageat aggeettggg 4260 aagtgeettt tegtaggeeg eggeateget ggageggegg tggeageat aggeettggg 4260 atteteetee etggegeet tggeagegg tggeagegge getegtggee 4320 tttaaggtea tggagegegg gatgeettee accgaggae tggeagege getegtgeege 4440 ggagegeegge gggaggggge tgtgeagtgg atgaacegge tggtaacet eggtageege 4440 ggagegeegge teteceegegg gtegtggggaegggggggggggggggggggggg	gcgatagcat	cactgatggc	attcacagec	tctatcacca	acccactcac	cacccaacat	
adgtgcttg tggatattt ggcaggtat ggagcaggtg tggcaggag aggccttggg 4260 tttaaggtac tggatatttt ggcaggtat ggagagggg tggcaggagg ctctctgggc 4320 tttaaggtac tggatagtgc tggatagtat ggagaggag tggtaact actccttgct 4380 atcctctccc ctggcgccct agtcgtggg gtcgtgtgcg cagcgatact gggtagaca 4440 gtgggcccag ggagagggg tgtgcagtgg gtcgtgtgcg cagcgatact gggtagaca 4440 gtgggcccag gggagagggg tgtgcagtgg atgaccgc cagcgatact gggtagaca 440 gtgggcccag gggagagggc tgtgcagtgg atgacaggag acgctgcagc acgttcacct 4560 ggataccact ctcatcactcac ctagtcttac catcactcac ctggtgaaga acgctgcaga acgtgtcaact 4620 gaggatattga ctcagtctac accggcatg tccaggcact caggcaga acgtgtcaaca 4620 ggctgttga ctgattcaa acgtgggtac acgggtataga atgtttgga ttggatatatg 4680 acggtgttga ctgattcaa acgtgggtac aagggagtct tccatcagg gtccattg gaccatggg accatgggat tccatcagggg ttcgatggag atcggtgggg tccattg gaccaggat acggtgtgggg ttcaacagg accatgggg accatgggat accatgggg accatgggggc cagcattatt tcaatggtggg ctcatggaga accatggaggat tccattagagg accatggaga accatggaggat accatgggg gttacactggag accatggagaggt accatgggg accatggaga accatggagagggt accatggagagaggagag							
aaggtgcttg tggatatttt ggacagttat gagacaggg tggcaggcg gtcgtggcc 4320 tttaaggtca tgagcggag gatgcctcc accgaggacc tggttaacct actcctgct 4380 atcctctccc ctggcgccct agtcgtcggg gtcgtgtgcg cagcgatact gcgtcggaca 4440 gtgggcccag gggaggggc tgtgcagtgg atgaaccggc tgatagcgtt cgcttcgcgg 4500 ggtaaccacg tctccccac cactattg cctgagagcg acgctgcagc acgtgtcact 4560 cagatctct ctagtcttac catcactcag ctgctgaaga acgttcacca gtggatcaac 4620 gaggactgct ccacgccatg ctccggctcg tggctaaaga atgtttggga ttggatatgc 4680 acggtgttga ctgatttcaa gacctggtc cagtcaaga tccttgcgga 4740 gtcccttct tctcatgtca acgtggtac aagggagtct ggcgggggg cggattgcgg atgcacact 4800 caaaccacct gcccattgtg agcacagatc accggacatt ggcgggggga cgcatcatt 4800 acggtgtggg cccatgtgg agcacagatc accggacatt ggcgggggg cgcatcatt 4800 accacggggc ctgccattgtg agcacagatc accggacatt tgaaaaacgg ttccatgagg 4860 atcgtgggg cccatgtgg agtacaccag gttgaggggtt tccatgagg gttgggggg agtacgtgggggggggg							
atcetetece etgegece agtegeteg getegtege cagegatect getegede 4440 getegeced getegede 4440 getegeced getegede 4440 getegeced getegede 4500 getagecede 4440 getegecede 4500 getagecede 4500 getag	aaggtgcttg	tggatatttt	ggcaggttat	ggagcagggg	tggcaggcgc	actcataacc	
atcetctecc ctggcgcct agtgcgggg gtcgttgtgcg cagcgatact gcgtcggcac 4440 gtgggcccag gggaggggg ttgtcactggg atgaccagcg tctccccac cagatcctct ctccccac cactactcag ctgtcaagag accttgcag acgtttcacc 4560 cagatcctct ctagtcttac catcactcag ctgctaagag ggcttcacca gtggatcaac 4620 gaggactgct ccacgccatg ctccggctcg tggctaagag atgtttgga ttggatatgc 4680 acggtgttga ctgatttcaa gacctggctc cagtccaaga ttgtttagga ttggatcaac 4620 gaggactgttga ctgatttcaa gacctggctc cagtccaagag atgtttgga ttggatatgc 4680 acggtgttga ctgatttcaa gacctggctc cagtccaagac tcctgcgcg attgccggga 4740 gtccccttct tctcatgtca acgtgggtac aacgggaggtct ggcggggga cggcatcatg 4800 acggtggggg ctaggagc ctaggacctg tagtaacacg ttgaaaaacgg ttccatgagg 4860 accacgggcc ctaggacctg tagtaacacg ttgaaaaacgg ttccatgagg 4860 gctgtgtgggg ctaggacctg tagtaacacg ttgaaaaacgg ttccatgagg 4860 gctgtgtgggg agtacgtgga ggttacgcgg ccaaattatt ctagggggct ttaacgcggt 4920 accactgga agtacgtgga ggttacgcgg gttggggggatt tccactacgt gagggggggggg	tttaaggtca	tgagcggcga	gatgccctcc	accgaggacc	tggttaacct	actccctqct	
gutageccag ggaaggggg tetecaca geactaty cetyagageg acgetycage acgttecact 4560 cagatectet ctagetetac catagtg cetyagageg acgetycage acgtycact 4560 gagagecety ctagetycaca geactaty cetyagagageg acgetycage acgtycact 4560 cagatectet ctagetetac catagety tetyagagagegagegagggaggaggaggaggaggaggaggagg	atcctctccc	ctggcgccct	agtcgtcggg	gtcgtgtgcg	cagcgatact	gcgtcggcac	
ggtaaccacg tctccccac gcactatgtg cctgaagacg acgttgcagc acgtgtact tagatcctct ctagtcttac catcactcag ttgctaagag ggcttcacc gtggatcac 4620 gaggactgct ccacgccatg ctccggctcg tggctaagag aggttttggg ttggatatacc 4680 acggtgttga ctgatttcaa gacctggctc cagtccaagc tcctgccgcg attgcaggg 4740 gtcccttct tctcatgtca acgtgggtac acggacatg tcctagggac cggcatcatg 4800 caaaccacct gcccatgtgg agcacagatc accggacatg tggaaaacagg ttcctagggg cggcatcatg 4800 acggtgtgggg ctaggaggcc ctaggacacg ttgaaaaacagg ttccatgagg 4920 accacgggcc ctccccggcg ctccccggcg ccaaattatt ctagggggct tccatgagg 4980 gctgctgagg agtacgtgg ggttacgcgg gttggggggat tccactacgg gagggggggggg	gtgggcccag	gggaggggc	tgtgcagtgg	atgaaccggc	tgatagcgtt	cgcttcgcgg	4500
gaggactgct ccacgccatg ctccggctc aggctaagag atgtttggga ttggatatgc 4680 acggtgttga ctgatttcaa gacctagctc cagtccaagc tcctgccgcg attgccagga 4740 gtcccttct tctcatgtca acgtgggtac aaggcaggtct ggcgggggga cggcatcatg 4800 caaaccacct gcccatgtgg agcacagatc acggacatg tggaaaaacgg ttccatgagg 4860 atcgtggggc catgagccc ctcccggcg ccaaattatt ctaggggcgc gtggggggt 4980 gctgctgtaag agtacgtga ggttacggg gtggggggtt tccactacgt gacggggtg 4980 gctgctgtaag agtacgtga ggttacggg gtggggggtt tccactacgt gacgggatg 5040 accactgaca acgtaaagtg cccgtgtcaag gttccggcc ccgaattatt cacagaagtg 5000 gatgggggtg ggttgcaag gttccggcc ccgaattatt cacagaagtg 5000 gatgggggtg gggttgcaa acgtcactacg ggagggggtt tccactacg ggaggaggtc 5000 gatgggggtg caggactgaa cacttcctgg tcggacgtaa cacttcctag ggaggaggtc acattcctgg tcggacgtaa cactacctg gttgggtcaa accttcctacg ggaggaggtc 5160 acattcctgg tcggacgtaa tcaatacctg gttgggtcaa agctcccaat ggaggagaggt 5220 ccgaatgtat aggctggcaa aggtctccc cccacattac ggcggagagg 5220 ggaagaggta aggctggcc acactcct gtgggggaacactcac cccacattac ggcggagagg 5220 ccattcatgag aggtagtaat tttgagactct ttcgaggccg gagatggcg ggaacatcac ccggtggag 5340 ctcatcatgag ccaacctcct gtggcggaag gagatgggg ggaacatcac ccgcgtggag 5340 ctcatcatgag aggaagtat ccgttccgag ggagatcctg cgagggtca aggaagagag 5520 agggaagtat ccgttccgg ggagatcctg cgagggtca ggaaattccc tcgagcgag ggaagatgag 5520 agggaagata ccgttccgg ggagatcctg ccactgttag agtctggaa ggaccggac 5640 tacgtccctc cagtggtaca cagggtgtca ttgcacacct ccactgttag agtctggaa ggaccggac 5640 tacgtccctc cagtggtaca caaagacct ccggcagcc ggaagatccg caaagacct tcggacgac ggaagagcc ggaagagcc ggaagagcc ggaaccgaa cacggccct tccaacgg ggagacccga ccactcctaggg ggagacccga ccactcctaggg ggaggaccga ggaggagccg ggaggagcg ggaggaccga accggacccga accggacccga accggacccca aggagagaccca aggagagacc aggagagacca aggagaacca aggagagacca aggagaacca aggagagacca aggagagacca aggagagacca aggagaacca aggaggaacca aggagaacca aggagaacca aggagaacca aggagaacca aggagaacca ag	ggtaaccacg	tctcccccac	gcactatgtg	cctgagagcg	acgctgcagc	acgtgtcact	4560
acggtgttga ctgatttca gacctggct cagtccaage tectgeege attgeegga 4740 gteectett teteatgte aegtgggtae aagggagete ggeggggega eggeateatg 4800 aegtggggge eecaatgg agcacagate aeggacatg tgaaaaaegg tecetgagg 4860 ategtgggge etaggacetg tagtaacaeg tggeatggaa catteecat taaeggetae 4920 getgetgagg agtacgtgg ggttaegegg gtgggggatt ecaatacgt gaeggggggg aecaacatgaca acgtaaaagtg ggttaegegg gtgggggatt tecactaegt gaeggggtg 4980 gatggggtg ggttgeaca accaetgaca acgtaaaagtg gtteeggee ecaaattatt ecacagaggg gaeggggggggatggatggatggatggatggatggatgga	cagatcctct	ctagtcttac	catcactcag	ctgctgaaga	ggcttcacca	gtggatcaac	4620
stccctttt tctcatgtca acgtggtac aagggagtct ggcgggcga cggcatcatg 4800 caaaccacct gcccatgtgg agcacagatc accggacatg tgaaaaacgg ttccatgagg 4860 atcgtgggcc ctaggacctg tagtaacacg tggcatggaa cattcccat taacgcgtac 4920 accacgggcc ctccccggcg ccaaattatt ctagggcgct gtggcgggtg 4980 gctgctgaag agtacgtgga ggttacgcgg gtgggggatt tccatcacgt gacgggcatg 5040 accactgaca acgtaaagtg cccgtgtcag gttccggccc ccgaattctt cacagaagtg 5100 gatgggggtg ggttgcacag gtacgccca gcgtgcaac ccctcacag ggaggaggtc 5160 acattcctgg tcgggctcaa tcaatacctg gttgggtcac agctccatg ggaggaggtc 5160 gctaagcgta ggctgccaa tccatacctg gttgggtcac agctccata ggcggagagcg 5220 ccggacgtag ggctgccaa tccatacctg gttgggtcac accgaccct cccacattac ggcggagagcg 5280 gctaagcgta ggctgccaa ggatctccc ccctctttgg ccagctcatc agctgaccag 5340 ctgtctgcgc cttccttgaa ggcaacatgc actacccgtc atgaccaccc ggagggagagca 5340 ctgtctgcgc cttccttgaa ggcaacatgc actacccgtc atgaccacc ggagggagagag 5520 agggaagtat ccgttccggc ggagatcctg ggagatcggg ggaacatcac ccgcgtggag 5460 tcagaaaata aggtagtaat tttggactct ttcgagccgc tccaagcgga ggagggatgag 5520 agggaagtat ccgttccggc ggagatcctg ggaggtcca ggaacttccc tcgagcgatg 5580 cccatatggg cacgcccgaa ttacaaccct ccactgttag agtcctggaa ggaccggaac 5580 cccatatggg cacgccgaa ttacaaccct ccacgtgtag aggacccggac 5580 cccatatgg cacgcccga ttacaaccct ccacgtgtag aggacccggac 5580 cccatatgg ggaagaggac ggggtgtcca ttgccgccg ccaaggcccc tcgaatacca 5700 cctccacgga ggaagagac gggttgtcct tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggacgtcc gaatcgtcgg ccaaggccc tcagaggca 5820 agggctcccatcatgc cccccttga ggggagccg ggggatccgg gatccaacac cgggtcttgg 5940 tctaccgtaa gcgagaggc tagtgagac ggggatcceg ggggatccca caatgcactg 6060 acaggcacct ttgctcgca agaaggccc caaaacctt ggcgcgaa gtcgtctac caacacttg caacactcc caacacttg gcgaaggcc tagtgacca ctaccggaac 6120 ctcgcgcaag agaaggtca agaaggtcc ctttgacaga cgtcgtcca acagttacc ctaccgggac 6180 gtgtccaag agaaggtca ctttgacaga caacttct acacactcc ctaccgggac 6180 gtgtccaagg agaaggtca agaaggtcc acaggtcca acagttacac ctaccgggac 6180 gtgtccaagg agaaggtcac ctttgacaga agaaggt	gaggactgct	ccacgccatg	ctccggctcg	tggctaagag	atgtttggga	ttggatatgc	4680
caaaccacct gcccatgtgg agcacagatc accggacatg tgaaaaacgg ttccatgagg 4860 atcgtggggc ctaggacctg tagtaacacg tggcatgga cattccccat taacgcgtac 4920 accacgggcc ctcccccggcg ccaaattatt tgaggcgct gtggcgggtg 4980 gctgctgagg agtacgtgag ggttacgcgg gtgggggatt tccactacgt gacggcgatg 5040 accactgaca acgtaaagtg cccgtgtcag gttccggcc cgaattctt cacagaagtg 5100 gatgggggtg tcggggggatg tccatcctgg ggaggagggcc cacattcctgg tcgggccaa tcaatacctg gttgggtcaa accattcctgg tcgggccaa tccatacgtg gttggggtcaa accattcctgg tcgggccaa tccatacgtg gttggggtcaa accattcctgg ggaggacgtaacccgaa cactcctaagggccaa ggctggccaa ggctgcaaacctcc tccatacgt gaggagaggc 5160 accatcgaggacgtag ggagtgccaa accatcct gggggtgcaa accacctcc accaacatca ggcggaagacg 5280 gctgctgaaggaagaata aggtagacaatcac ccctccttgg caagctcata aggtagacaa 5280 gctgctgaaaaata aggtagtaaa ttttggactct ttcgagccgc accactaca ggaagagagacg 5460 tcaaaaaaaa aggtagtaaa ttttggactct ttcgagccgc tccaaagcgga ggaagatgag 5520 agggaagata ccgttccgg aggaatcctg ggaagatcct ggaagatcca caacctcggaagagacgaagacccaaccccaatacgggaagataa cacccccaatacgggaagataa cacccccaatacgggaagataa ccggtggaagataa ccgggtgtcaa tttggagccg ggaaaatccc tcgagcgatga 5580 accatatagg caagaagaaca cacccccaacgaga ggaagatgag 5520 agggaagataa ccgttccaacgagaagacgaacaccccaacgagaagacgaacaccccaacgagaaga	acggtgttga	ctgatttcaa	gacctggctc	cagtccaagc	tcctgccgcg	attgccggga	
accacgggcc ctaggacctg tagtaacacg tggcatggaa cattccccat taacgcgtac 4920 accacgggcc cttgcacgcc ctccccggcg ccaaattatt ctaagggcgt tgggcgggtg 4980 gctgctgagg agtacgtgga ggttacgggg gtgggggatt tccactacgt gacgggcatg 5040 accactgaca acgtaaagtg cccgtgtcag gttccggcc ccgaattctt cacagaagtg 5100 gatgggggtc gggggtcaa tcaatacctg gtgggggatc cccactacg ggaggaggtc 5160 acattcctgg tcgggctcaa tcaatacctg gttgggtcac agctccatac ggaggaggcc gcaaactgacgcccccccccc	gtccccttct	tctcatgtca	acgtgggtac	aagggagtct	ggcggggcga	cggcatcatg	
accacgage citececage citececage caaattatt ctaggeget gtageggtg 4980 gctgctgaag agtacgtga ggttacgeg gtgggggatt tecactacgt gacggcatg 5040 accactgaca acgtaaatg cccgtgtcag gttccggcc ccgaattctt cacagaagtg 5100 gatggggtg ggttgcacag gttccggccc ccgaattctt cacagaagtg 5100 acattectg tcgggctcaa tcaatacctg gttgggtcac agcteccatg ggaggaggtc 5160 acattectgg tcgggctcaa tcaatacctg gttgggtcac agcteccatg ggaggaggcc 5220 ccggacgtag ggctgccag ggatetece ccctcttgg ccacacttac ggcggagacg 5220 gctaagcgta ggctgccag gggatetece ccctccttgg ccacactatac ggcggagacg 5280 gctaagcgta ggctggcag gggatetece accacactac atgactecec ggacgctgac 5400 ctcatcgagg ccaacctcct gtgggggaa gaggatggg ggaacatcac ccgcgtggag 5460 tcatcagagg ccaacctcct gtggcggcag gagatgggcg ggaacatcac ccgcgtggag 5520 agggaagtat ccgttccggc ggagatectg cggaggtca ggaaattccc tcgaagaata aggtagtaat tttggacct ttcgagccgc tcaaagcgga ggaggatgag 5520 agggaagtat ccgttccgg ggagatcctg cggaggtca aggcccggac 5640 tacgtcctc cagtggtaca cgggtgtca ttgcacacct ccactgttag agtcctggaa ggacccggac 5640 tacgtccctc cagtggtaca cgggtgtcca ttgccgcctg ccaaggccc tccgaatacca 5700 cctccacagg ggaagaagaa ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagctc gaatcgtcgg ccgtcgacag cggcacggca	caaaccacct	gcccatgtgg	agcacagatc	accggacatg	tgaaaaacgg	ttccatgagg	
gctgctgagg agtacgtgga ggttacgcgg gtgggggatt tccactacgt gacgggcatg 5040 accactgaca acgtaaagtg cccgtgtcag gttccggccc ccgaattctt cacagaagtg 5100 gatgggggtc ggttgcacag gtacgctcca gcgtgcaaac ccctcctacg ggaggaggtc 5160 acattcctgg tcgggctcaa tcaatacctg gttgggtcaa acgtcccatg ggaggaggtc 5160 acattcctgg tcgggctcaa tcaatacctg gttgggtcac acctcctacg ggaggaggcc 5220 acgacgtag cagtgctcac tccatgctc accgacccct cccacattac ggcggagacg 5220 gctaagcgta ggctggccag gggatctccc ccctcttgg ccacacttac agctacgcag 5340 actgctctgcg cttccttgaa ggcaacatgc actacccgtc atgactccc ggacgctgac 5400 acgacaatacactccgggagagatactccg aggagatactcc aggaggatacg 5340 agggaagtaa aggtagtaat atttggactct ttcgaggcgc tccaagggag ggaacatcac aggaggatgag 5520 agggaagtat acggtccgg ggagatcctg aggaggtcaa ggacgggag 5520 agggaagtat acggtcccga ttacaaccct agggggtcaa aggacgggaactccg aggaaattccc aggaggatcg 5580 agggaagtatacggg aggagatccg aggaaattccc aggaggatcg 5580 agggaagtacccg aggaggtcaa aggacccggaa ggaacccggaa ggacccggaa 5580 acccatatggg aggaagagac ggttgtcca ttgcgcgctg aggacccggaa 5580 acccacactgg aggaagaggac ggttgtcca ttgcgcgctg aggacccggaa 5500 acccacactga ggaagaggac ggttgtcca ttgcgcgctg aggacccgaagac 5640 accacactgg aggaagagac ggttgtcct atgagaactac acctggaaga ggacccggaa 5580 acccacactga agacgccc acaagacctc aggagagaccg ggagaaccac accacactga aggaccacgac 55700 acccacacga ggaagaggac ggttgtcct gaatcgcga aggacacgac 5580 accggcccca acaagacctc acaagacctc aggagagaccg ggggaacccg atccacacactg aggacacgac 5580 accggccccacacactccacacactg ggggaacccg atccacacactg accacacactg 55700 accggcccccacacacactc aggagagaccg ggggaaccca atccacacacactg accggcacacacacacacacacacacacacacacacac	accycygygc	ctaggacctg	tagtaacacg	tggcatggaa	cattccccat	taacgcgtac	
accactgaca acgtaaagtg cccgtgtcag gttccggccc ccgaattett cacagaagtg 5100 gatggggtgc ggttgcacag gtacgetca gcgtgcaaac ccctcetacg ggaggaggtc 5160 acattcetg tcgggctcaa tcaatacetg gttgggtcac agctcccatg cgagcccgaa 5220 ccggacgtag gatggctcac ttccatgetc accgaccet cccacattac ggcggagacg getaagcgtag cagtgctcac gttccatgetc accgaccet cccacattac ggcggagacg 5280 gctgtctac gctgccag gggatetccc ccctcettgg ccagetcatc agctagccag 5340 ctgtctgcgc cttccttgaa ggcaacatga gagatgggcg ggaacatcac ccgcgtggag 5460 tcaatcgagg ccaacetcet gtgggcgcag gagatgggcg ggaacatcac ccgcgtggag 5460 tcagaaaata aggtagtaat tttggactet ttcgagccgc tccaagcgga ggaggatgag 5520 agggaagtat ccgttccgg ggagatcctg cggaggtcca ggaaattccc tcgagcgatg 5580 cccatatggg cacgccgaa ttacaaccct ccactgttag agtcctggaa ggacccggac 5640 tacgtccct cagtggtaca cgggtgtcca ttgccgcctg ccaaggcccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcctg tcagaactca ccggtgtttc tgccttggcg 5760 gagctcgca caaagacctt cggcagctc gaatcgtcgg cgtcgacag cggcacggca							
gatggggtgc ggttgcacag gtacgctcca gcgtgcaaac ccctcctacg ggaggaggtc 5160 ccggacgtag tcggcccaa tcaatacctg gttgggtcac agctcccatg cgagcccgaa 5220 ccggacgtag cagtgctcac ttccatgctc accgacccct cccacattac ggcggagacg 5280 gctaagcgta ggctggccaa gggatctccc ccctccttgg ccagctcatc agctagccag 5340 ctgtctgcgc cttccttgaa ggcaacatgc actacccgtc atgactcccc ggacgctgac 5400 ctactcatcgagg ccaacctcct gtggcggcag gagatgggcg ggaacatcac ccgcgtggag 5460 ctcatcgagg ccaacctcct ttcgacgcgc gagagtccag ggaggatgag 5520 agggaagata ccgttccggc ggagatcctg cggaggtcca ggaaattccc tcgagcgatg 5580 cccatatggg cacgccgga ttacaaccct ccactgttag agtcctggaa ggacccggac 5640 cccatatggg cacgcccga ttacaaccct ccactgttag agtcctggaa ggacccggac 5640 cccatatggg cacgcccga ttacaaccct ccactgttag agtcctggaa ggacccggac 5640 cccatatggg ggaagaggac ggttgtcca ttgccgcctg ccaaggccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcct ttgcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca							
ccggacgtag cagtgctcac ttccatgctc accgacccct cccacattac ggcggagacg 5280 gctaagcgta ggctggcag gggatctccc ccctccttgg ccagctcatc agctagccag 5340 ctgtctgcgc cttccttgaa ggcaacatgc actacccgt atgactcccc ggacgctgac 5400 ctatctgagg ccaacctcct gtggcggcag gagatgggcg gagatgggcg gagacatcac ccgcgtggag 5460 tcatcgagg ccaacctcct gtggcggcag gagatgggcg ggaacatcac ccgcgtggag 5520 aggaagatat aggtagtaat tttggactct ttcgagccgc tccaagcgga ggaggatgag 5520 agggaagatat ccgttccggc ggagatcctg cggaggtcca ggaaattccc tcgagcgatg 5580 cccatatggg cacgccgga ttacaaccct ccactgttag agtcctggaa ggaccggatg 5580 cccatatggg cacgccgga ttacaaccct ccactgttag agtcctggaa ggacccggac 5640 tacgtccctc cagtggtaca cgggtgtcca ttgccgcctg ccaaggccc tccgatacca 5700 cctccacgga ggaagagac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca	gatggggtgg	acytaaayty	atacactaca	gratagaaa	ccgaattctt	cacagaagtg	
Ccggacgtagcagtgctcacttccatgctcaccgaccctcccacattacggcggagacg5280gctaagcgtaggctggccaggggatctcccccctccttggccagctcatcagctagccag5340ctgtctgcgccttccttgaaggcaacatgcactacccgtcatgactccccggacgctgac5400ctcatcgaggccaacctcctgtggcggcaggagatgggcgggaacatcacccgcgtggag5460tcaagaaaataaggtagtattttggactctttcgagccgctcaagacggaggagatgatgag5520ccaatatgggcagtcccggattacaaccctccaactgttagagacccggac5640ccatatgggcagtggtacacgggtgtccattgccgcctgccaaggccctccgataccacctccacggaggaagaggacggttgtcctttgccgcctgccaaggccctccgataccacctccacggaggaagagccggcgacgcgggatccgacagcggcacggagagctcgccacaaagaccttcggcagctccgaatcgtcggcgtcgacagc5820acggcctctcctgaccagccctccgacgacggcgacgcgggatccgacgt5820acggcctctcctgaccagccggggagccgggggatcccgatctacgca5880tcctccatgcccccctttgagggggagccgggggatcccgatctacgca5940tctaccataggcgagaggcgatgaaaccacaacactgcctacacactg6000acaggcgcctgatcacgcatgcgctgcggatgatgccacaacactctcaacactctcaacactctctgcggcagaagaaggtcacctttgacagactgcaggtccctacacggaccctaccggacc6120ctgcggcagagaaggtcac							
gctaagcgta ggctggcag gggatctccc ccctcttgg ccagctcatc agctagcag 5340 ctgtctgcg cttccttgaa ggcaacatgc actacccgtc atgactcccc ggacgctgac 5400 ctcatcgagg ccaacctcct gtggcggag gagatgggcg ggaacatcac ccgcgtggag 5460 tcagaaaata aggtagtaat tttggactct ttcgagccgc tccaagcgga ggaggatgag 5520 agggaagtat ccgttccggc ggagatcctg cggaggtccc ggaaattccc tcgagcgatg 5580 cccatatggg cacgcccga ttacaacct ccactgttag agtcctggaa ggaccggac 5640 tacgtcctc cagtggtaca cggggtcca ttgccgcctg ccaaggccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagtcc gaatcgtcgg ccgtcgacag cggcacggca							
ctgtctgcgc cttccttgaa ggcaacatgc actacccgtc atgactcccc ggacgctgac 5400 ctcatcgagg ccaacctcct gtggcggcag gagatgggcg ggaacatcac ccgcgtggag 5460 tcagaaaata aggtagtaat tttggactct ttcgagccgc tccaagcgga ggaggatgag 5520 agggaagtat ccgttccggc ggagatcctg cggaggtcca ggaaattccc tcgagcgatg 5580 cccatatggg cacgcccgga ttacaacct ccactgttag agtcctggaa ggacccggac 5640 tacgtccctc cagtggtaca cgggtgtcca ttgccgcctg ccaaggcccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca							
Ctcatcgagg ccaacctcct gtggcggcag gagatgggcg ggaacatcac ccgcgtggag 5460 tcagaaaata aggtagtaat tttggactct ttcgagccgc tccaagcgga ggaggatgag 5520 agggaagtat ccgttccggc ggagatcctg cggaggtcca ggaaattccc tcgagcgatg 5580 cccatatggg cacgccgga ttacaacct ccactgttag agtcctggaa ggacccggac 5640 tacgtcctc cagtggtaca cgggtgtcca ttgccgcctg ccaaggcccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca							
tcagaaaata aggtagtaat tttggactct ttcgagccgc tccaagcgga ggaggatgag 5520 agggaagtat ccgttccggc ggagatcctg cggaggtcca ggaaattccc tcgagcgatg 5580 cccatatggg cacgcccgga ttacaaccct ccactgttag agtcctggaa ggacccggac 5640 tacgtccctc cagtggtaca cgggtgtcca ttgccgcctg ccaaggcccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca	ctcatcgagg	ccaacctcct	ataacaacaa	gagatgggcg	ggaacatcac	ccacataaaa	
agggaagtat ccgttccggc ggagatcctg cggaggtcca ggaaattccc tcgagcgatg 5580 cccatatggg cacgccgga ttacaacct ccactgttag agtcctggaa ggacccggac 5640 tacgtcctc cagtggtaca cgggtgtcca ttgccgctg ccaaggcccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca	tcagaaaata	aggtagtaat	tttggactct	ttcgagccgc	tccaaqcqqa	ggaggatgag	
CCCatatggg cacgccgga ttacaacct ccactgttag agtcctggaa ggacccggac 5640 tacgtccctc cagtggtaca cgggtgtcca ttgccgctg ccaaggcccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca	agggaagtat	ccgttccggc	ggagatcctg	cggaggtcca	ggaaattccc	tcgagcgatg	
tacgtccctc cagtggtaca cgggtgtcca ttgccgcctg ccaaggcccc tccgatacca 5700 cctccacgga ggaagaggac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca	cccatatggg	cacgcccgga	ttacaaccct	ccactgttag	agtcctggaa	ggacccggac	
CCtccacgga ggaagagac ggttgtcctg tcagaatcta ccgtgtcttc tgccttggcg 5760 gagctcgcca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca	tacgtccctc	cagtggtaca	cgggtgtcca	ttgccgcctg	ccaaggcccc	tccgatacca	5700
gagctcgcca caaagacctt cggcagctcc gaatcgtcgg ccgtcgacag cggcacggca	cctccacgga	ggaagaggac	ggttgtcctg	tcagaatcta	ccgtgtcttc	tgccttggcg	5760
acggcetete etgaceagee etecgacgae ggegaegegg gateegaegt tgagtegtae 5880 teetecatge ecceettga gggggageeg ggggateeeg ateteagega egggtettgg 5940 tetacegtaa gegaggagge tagtgaggae gtegtetget getegatgte etacacatgg 6000 acaggegeee tgateacgee atgegetgeg gaggaaacea agetgeeeat caatgeactg 6060 ageaactett tgeteegtea ecacaacttg gtetatgeta eaacateteg eagegeaage 6120 etgeggeaga agaaggteae etttgacaga etgeaggtee tggaegaeca etacegggae 6180 gtgeteaagg agatgaagge gaaggegtee acagttaagg etaaacttet ateegtggag 6240	gagctcgcca	caaagacctt	cggcagctcc	gaatcgtcgg	ccgtcgacag	cggcacggca	5820
tectecatge dececttga gggggageeg ggggateeeg ateteagega egggtettgg 5940 tetacegtaa gegaggage tagtgaggae gtegtetget getegatgte etacacatgg 6000 acaggegeece tgateacgee atgegetgeg gaggaaacea agetgeecat caatgeactg 6060 ageaactett tgeteegtea ecacaacttg gtetatgeta caacateteg eagegeaage 6120 etgeggeaga agaaggteac etttgacaga etgeaggtee tggaegacea etacegggae 6180 gtgeteaagg agatgaagge gaaggegtee acagttaagg etaaacttet ateegtggag 6240	acggcctctc	ctgaccagcc	ctccgacgac	ggcgacgcgg	gatccgacgt	tgagtcgtac	5880
acaggegece tgateaegee atgegetgeg gaggaaacea agetgeecat caatgeaetg 6060 ageaactett tgeteegtea ceacaacttg gtetatgeta caacateteg cagegeaage 6120 ctgeggeaga agaaggteae etttgacaga etgeaggtee tggaegaeca etacegggae 6180 gtgeteaagg agatgaagge gaaggegtee acagttaagg etaaacttet atcegtggag 6240	tcctccatgc	cccccttga	gggggagccg	ggggatcccg	atctcagcga	cgggtcttgg	
agcaactett tgeteegtea ceacaacttg gtetatgeta caacateteg cagegeaage 6120 Ctgeggeaga agaaggteac etttgacaga etgeaggtee tggaegaeea etacegggae 6180 gtgeteaagg agatgaagge gaaggegtee acagttaagg etaaacttet atcegtggag 6240	tctaccgtaa	gcgaggaggc	tagtgaggac	gtcgtctgct	gctcgatgtc	ctacacatgg	
Ctgcggcaga agaaggtcac ctttgacaga ctgcaggtcc tggacgacca ctaccgggac 6180 gtgctcaagg agatgaaggc gaaggcgtcc acagttaagg ctaaacttct atccgtggag 6240	acaggcgccc	tgatcacgcc	atgcgctgcg	gaggaaacca	agctgcccat	caatgcactg	
gtgeteaagg agatgaagge gaaggegtee acagttaagg etaaaettet ateegtggag 6240	agcaactctt	tgctccgtca	ccacaacttg	gtctatgcta	caacatctcg	cagcgcaagc	
yaayuuuyta aguugauguu cocacattog gocagatota aatttggota tggggcaaag 6300							
	yaaycctgta	agergaegee	eccacattcg	gccagatcta	aatttggcta	cggggcaaag	6300

gacgtccgga	acctatccag	r caaggccgtt	: aaccacatco	gctccgtgtg	gaaggacttg	6360
					ggttttctgc	6420
					agatttgggg	6480
gttcgtgtgt	: gcgagaaaat	ggccctttac	gatgtggtct	ccaccctccc	: tcaggccgtg	6540
					cctggtgaat	6600
					ttttgactca	6660
					tgacttggcc	6720
cccgaagcca	gacaggccat	aaggtcgctc	: acagagcggc	: tttacatcgg	gggccccctg	6780
actaattcta	aagggcagaa	ctgcggctat	: cgccggtgcc	: gcgcgagcgg	tgtactgacg	6840
accagetgeg	gtaataccct	cacatgttac	ttgaaggccg	ctgcggcctg	tcgagctgcg	6900
					ctgtgaaagc	6960
					gactagatac	7020
					aacatcatgc	7080
					tctcacccgt	7140
					tccagtcaat	7200
actcatttat	gcaacatcat	tatagataga	cecaeettgt	gggcaaggat	gatcctgatg	7260
acceattict	cotoctatect	cctageteag	gaacaacttg	aaaaagccct	agattgtcag	7320
categogy	cougulatio	cattgageca	cttgacctac	ctcagatcat	tcaacgactc	7380
teatgectea	gegeattice	actecatage	tactctccag	gcgagaccaa	tagggtgget	7440
gtacacacta	ggaaacttgg	ggtaccgccc	ttgcgagtct	ggagacateg	ggccagaagt	7500
aactggggca	tacage	ccaggggggg	agggctgcca	cttgtggcaa	gracerette	7560
ttatccacct	cattoottaa	teattagage	actccaatcc	cggctgcgtc	ccagttggat	7620
accedacec	ggttegttge	ataataasta	gggggagaca	chatacacag	cetgtetegt	7680
ctactcccca	accoatosac	graggractaa	ctcctacttt	ctgtaggggt	aggeatetat	7740
tttccctttt	ttttttt	ttttttt	acactccagg ttttttttt	ttttttt	ttataattt	7800
tttttcctct	ttttttcctt	ttettteett	tggtggctcc	atettagee	tagtgaggg	7860
tanctutuaa	aggreeatga	accacttaec	tgcagagagt	accetageee	acatatataa	7920 7980
agatcaagta	cttctagaga	attetagett	ggcgtaatca	taataataa	tatttactat	8040
gtgaaattgt	tatcagetea	caattccaca	caacatacga	accadaaaca	tasaatataa	8100
agcctgggat	acctaataaa	tgagctaact	cacattagtt	acattacact	cactgcccgc	8160
tttccagtcg	ggaaacctgt	cataccaact	ccattagtga	atrotroaac	acacaaaaa	8220
aggcggtttg	cgtattgggc	gcacttccgc	ttcctcgctc	actgactcgc	tacactcatt	8280
cattcaacta	caacaaacca	tatcagetea	ctcaaaggcg	gtaatacggt	tatccacaga	8340
atcaggggat	aacgcaggaa	agaccatgtg	agcaaaaggc	cagcaaaaagg	ccaggaaccg	8400
taaaaaggcc	acattactaa	cqtttttcca	taggctccgc	cccctgacg	agcatcacaa	8460
aaatcgacgc	tcaagtcaga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcgtt	8520
tccccctgga	agctccctcg	tacactetee	tgttccgacc	ctaccactta	ccggatacct	8580
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	8640
cagttcggtg	taggtcgttc	gctccaagct	gggctgtgtg	cacqaacccc	ccqttcaqcc	8700
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcc	aacccggtaa	gacacgactt	8760
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcggtgc	8820
tacagagttc	ttgaagtggt	ggcctaacta	cggctacact	agaaggacag	tatttggtat	8880
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	8940
acaaaccacc	gctggtagcg	gtggttttt	tgtttgcaag	cagcagatta	cgcgcagaaa	9000
aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	9060
aaactcacgt	taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	9120
			ctaaagtata			9180
cagttaccaa	tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	9240
catagttgcc	tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	9300
ccccagtgct	gcaatgatac	cgcgagaacc	acgctcaccc	gcaccagatt	tatcagcaat	9360
aaaccagcca	gccggaagtg	cgctgcggag	aagtggtcct	gcaactttat	ccgcctccat	9420
			agtaagtagt			9480
taacgtcgtt	gccatagcaa	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	9540
attcagctcc	ggctcccaac	gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	9600
agcggttagc	teetteggte	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	9660
actcatggtt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	9720
ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	9780
			taccgcgcca			9840
gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca	aggatettae	cgctgttgag	9900
accoagttcg	acguaaccca	cccgtgcacc	caactgatct	tcagcatctt	ttactttcac	9960
cagogtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	10020

gacacggaaa tgttgaatac tcatactctt cc gggttattgt ctcatgagcg gatacatatt tg ggttccgcgc acatttcccc gaaaagtgcc ac gacattaacc tataaaaata ggcgtatcac ga tgacggtgaa aacctctgac acttgcagct cc ggatgccggg agcaggcaag cccgtcaggg cg ctggcttaac tatgcggcat cagagcagat tg aataccgcac agatgcgtaa ggagaaaata cc ccgcaactgt tgggaagggc ggtcagtacg cg agggggatgt gctgcaaggc gattaagttg gg ttgtaaaacg acagccaatg aattgaagct ta actcactata	aatgtatt tagaaaaata aacaaatagg 10140 ctgacgtc taagaaacca ttattaccat 10200 agccettt cgtctagcgc gtttcggtga 10260 cgcagacg gtcacagct gtcgtgagtactgagg gtacaccaga tgcggtgtga gcatcagc ctccattcgc cattcagact 10500 cttcttcg ctattacgcc aactggcgaa 10620 taacgcca gggttttccc aatcacgacg 10620
<210> 4 <211> 23 <212> DNA <213> Primer oligonucleotide	
<400> 4 acatgatctg cagagaggcc agt	23
<210> 5 <211> 26 <212> DNA <213> Primer oligonucleotide	23
<400> 5 gacasgctgt gatawatqtc tccccc	
<210> 6 <211> 21 <212> DNA <213> Primer oligonucleotide	26
<400> 6	
tggctctcct caagcgtatt c	. 21
<210> 7 <211> 23 <212> DNA <213> Primer oligonucleotide	
<400> 7 actetetgea gteaagegge tea	23
<210> 8 <211> 21 <212> DNA <213> Primer oligonucleotide	23
<400> 8 cagtggatga accggctgat a	21
<210> 9 <211> 23 <212> DNA <213> Primer oligonucleotide	
<400> 9 ggggcgacgg catcatgcaa acc	23

<210> 10 <211> 23 <212> DNA <213> Primer oligonucleotide	
<400> 10 caggacctgc agtctgtcaa agg	23
<210> 11 <211> 17 <212> DNA <213> Primer oligonucleotide	
<400> 11 cgggagagcc atagtgg	17
<210> 12 <211> 19 <212> DNA <213> Primer oligonucleotide	
<400> 12 agtaccacaa ggcctttcg	19
<210> 13 <211> 21 <212> DNA <213> Probe	
<400> 13 ctgcggaacc ggtgagtaca c	21

(19) World Intellectual Property Organization International Bureau





(43) International Publication Date 1 August 2002 (01.08.2002)

(10) International Publication Number WO 2002/059321 A3

C12N 15/40, (51) International Patent Classification7: 15/51, 5/10, C07K 14/18, C12N 15/85, 7/00, G01N 33/50

(IT). PAONESSA, Giacomo [IT/IT]; IRBM, Via Pontina

Patent Department, Terlings Park, Eastwick Road, Harlow,

(21) International Application Number:

PCT/EP2002/000526

(22) International Filing Date: 16 January 2002 (16.01.2002)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/263,479 23 January 2001 (23.01.2001)

(71) Applicant (for all designated States except US): ISTI-TUTO DI RICERCHE DI BIOLOGIA MOLECO-LARE P. ANGELETTI SPA [IT/IT]; Via Pontina Km. 30, 600, I-00040 Pomezia (Rome) (IT).

(72) Inventors; and

(75) Inventors/Applicants (for US only): DE FRANCESCO, Raffaele [IT/IT]; IRBM, Via Pontina Km. 30, 600, I-00040 Pomezia (Rome) (IT). MIGLIACCIO, Giovanni [IT/IT]; IRBM, Via Pontina Km. 30, 600, I-00040 Pomezia (Rome) Km. 30, 600, I-00040 Pomezia (Rome) (IT).

(74) Agent: THOMPSON, John; Merck & Co., Inc., European

(81) Designated States (national): CA, JP, US.

(84) Designated States (regional): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:

with international search report

Essex CM20 2QR (GB).

- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- (88) Date of publication of the international search report: 25 March 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



(54) Title: HEPATITIS C VIRUS REPLICONS AND REPLICON ENHANCED CELLS

(57) Abstract: The present invention features nucleic acid containing one or more adaptive mutations, and HCV replicon enhanced cells. Adaptive mutations are mutations that enhance HCV replicon activity. HCV replicon enhanced cells are cells having an increased ability to maintain an HCV replicon.